09-14-09

EXPRESS MAIL: EH960004675US

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re patent application of	:	Staehler et al.
Application No.	:	10/579,769
Title	:	Highly Parallel Template-Based DNA Synthesizer
Art Unit	:	1637
Examiner	:	Thomas, David C.

Pro Se
OK to Enter
mcs
9-28-09

Third Party Submission Under 37 CFR 1.99

Hon. Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

The undersigned submits the following five patents and publications, which are relevant to the above-referenced pending application. This submission complies with all of the requirements of 37 CFR 1.99. Accordingly, the undersigned respectfully requests that the following five patents and publications are entered in the above-referenced application:

- Nilsson et al., "Real-Time Monitoring of DNA Manipulations Using Biosensor Technology," Analaytical Biochemistry, vol. 224, pages 400-408, which published in 1995
- U.S. Pat. No. 5,795,714 to Cantor et al., which published Aug. 18, 1998
- Liu et al., "DNA Computing On Surfaces," Nature, vol. 403, pages 175-179, which published Jan. 13, 2000
- U.S. Pat. No. 6,534,271 to Furste et al., which published Mar. 18, 2003
- U.S. Pat. No. 6,632,641 to Brennan et al., which published Oct. 14, 2003

Upon learning of Applicant's recent amendments, a prior art search for relevant patents and publications was initiated. On August 21, 2009, upon identification of two relevant patents but while the search was still ongoing, the undersigned made a third party submission under 37 CFR 1.99. The submission was approved and entered on September 3, 2009. Now, upon identification of five additional patents and publications, the undersigned makes a supplemental

09/15/2009 MBELETE1 00000018 10579769
01 FC:1806 180.00 OP
02 FC:1464 130.00 OP

Upon learning of Applicant's recent amendments, a prior art search for relevant patents and publications was initiated. On August 21, 2009, upon identification of two relevant patents but while the search was still ongoing, the undersigned made a third party submission under 37 CFR 1.99. The submission was approved and entered on September 3, 2009. Now, upon identification of five additional patents and publications, the undersigned makes a supplemental submission under 37 CFR 1.99. The combined submission is in compliance with 37 CFR 1.99(d) because it is limited to less than ten total patents and publications.

This submission is permitted under 37 CFR 1.99(e) because they could not have been submitted to the Office earlier. See MPEP 1134.01(I). Applicants' amendments filed 07/28/2009 changed the scope of the claims to an extent that could not reasonably have been anticipated by a person reviewing the published application during the period specified in 37 CFR 1.99(e). The extent of the change in the scope of the amended claims is evidenced the Examiner's statement that "further search will be necessary in the next round of prosecution." See the Examiner Interview Summary dated 08/05/2009. Applicants also noted this point in their own Interview Summary, which stated that "Agreement was not reached and the Examiner stated that further search was necessary." See Applicants' Amendment and Response filed 07/28/2009 at page 10.

Copies of the five patents and publications are submitted herewith.

The fees set forth in 37 CFR 1.17(i) and (p) are also submitted herewith.

This submission was served upon the applicant in accordance with 37 CFR 1.248.

Respectfully,

Sandra A. Brockman-Lee
Reg. No. 44,045

EXPRESS MAIL: EH960004675US

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re patent application of	:	Staehler et al.
Application No.	:	10/579,769
Title	:	Highly Parallel Template-Based DNA Synthesizer
Art Unit	:	1637
Examiner	:	Thomas, David C.

Transmittal

Attached herewith are the following documents

- Third Party Submission Under 37 CFR 1.99
- Nilsson et al., "Real-Time Monitoring of DNA Manipulations Using Biosensor Technology," Analaytical Biochemistry, vol. 224, pages 400-408, which published in 1995
- U.S. Pat. No. 5,795,714 to Cantor et al., which published Aug. 18, 1998
- Liu et al., "DNA Computing On Surfaces," Nature, vol. 403, pages 175-179, which published Jan. 13, 2000
- U.S. Pat. No. 6,534,271 to Furste et al., which published Mar. 18, 2003
- U.S. Pat. No. 6,632,641 to Brennan et al., which published Oct. 14, 2003
- Authorization to charge required fees

Certificate of Service

I hereby certify on this 11th day of September 2009, that a true and correct copy of the foregoing "Third Party Submission Under 37 CFR 1.99" was mailed by first-class mail, postage paid, to:

Robert B. Murray, Attorney for the Applicants
ROTHWELL, FIGG, ERNST & MANBECK, P.C.
1425 K STREET, N.W., SUITE 800
WASHINGTON, DC 20005

Respectfully,

Sandra A. Brockman-Lee

ANALYTICAL BIOCHEMISTRY 224, 400–408 (1995)

Real-Time Monitoring of DNA Manipulations Using Biosensor Technology

Peter Nilsson,* Björn Persson,† Mathias Uhlén,* and Per-Åke Nygren*,[1]

*Department of Biochemistry and Biotechnology, Royal Institute of Technology (KTH), S-100 44 Stockholm, Sweden; and †Pharmacia Biosensor AB, S-751 82 Uppsala, Sweden

Received July 5, 1994

The potential of real-time biospecific interaction analysis technology for applications in molecular biology is described. DNA fragments are immobilized onto a biosensor surface using the high-affinity streptavidin–biotin system and subsequently used to monitor different unit operations in molecular biology, e.g., DNA strand separation, DNA hybridization kinetics, and enzymatic modifications. A model system comprising six oligonucleotides was used, which can be assembled into a 69-bp double-stranded DNA fragment. Using this system, the biosensor approach was employed to analyze multistep solid-phase gene assembly and the performance of different enzymes routinely used for the synthesis and manipulation of DNA. In addition, a concept for the determination of single-point mutations in DNA samples is described. © 1995 Academic Press, Inc.

Recent development in instrumentation for biospecific interaction analysis (BIA)[2] using biosensor technology has made it possible to monitor biological events in real-time (1). The biosensor used in this study is based on surface plasmon resonance (SPR) for detection of changes in refractive index over time at a sensor surface (1,2). These changes are proportional to the mass of molecules bound to the surface and are shown in a so-called sensorgram as resonance units (RU) over time. From the results, stoichiometric and kinetic data for the interaction can be determined (3). BIA has to date been used primarily for the study of different types of protein–protein interactions, e.g., antibody–antigen recognition (3), but also to some extent for DNA–protein interactions (4,5) and DNA–DNA hybridizations (6).

[1] To whom correspondence should be addressed. Fax: +46 8 245452.

[2] Abbreviations used: BIA, biospecific interaction analysis; SPR, surface plasmon resonance; RU, resonance units; SDS, sodium dodecyl sulfate; PCR, polymerase chain reaction.

In this work, we have investigated the extension of this technology into molecular biology applications for the analysis of basic DNA-related techniques such as hybridization and DNA strand separation as well as techniques including enzymatic action, such as ligation, endonuclease cleavage, and DNA synthesis. The potential of this tool is shown for the analysis of different types of such operations. In addition, a concept for real-time DNA minisequencing is described.

MATERIALS AND METHODS

Oligonucleotides

Six oligonucleotides, designed for assembly into a 69-bp fragment (Fig. 1), were synthesized on an automated DNA synthesizer (Gene Assembler Plus, Pharmacia Biotech) according to the manufacturer's recommendations. The oligonucleotides were purified using an FPLC pepRPC5/5 column (Pharmacia Biotech) (7). Phosphorylation was performed according to Sambrook *et al.* (8), where indicated. Oligonucleotides 1 and 6 were synthesized both with and without a 5′ biotin group using biotin-phosphoramidite (Clontech), for use in different experiments.

Biosensor Instrument

A BIAcore instrument (Pharmacia Biosensor) was used in all experiments. Sensor chips SA5 (research grade), precoated with approximately 4000 RU [1000 RU corresponds to approximately 1 ng/mm^2 (1)] of streptavidin, were from Pharmacia Biosensor. Before use in experiments, the sensor chips were treated with several (five to six) pulses of 50 mM sodium hydroxide to precondition the surface (corresponding to a decrease of approximately 500 RU). The experiments were performed at a temperature of 25°C. The flow rate was 2 μl/min, except for the cleavage experiment where 1 μl/min

0003-2697/95 $6.00

was used. The running buffer was 1× HBS [10 mM Hepes, pH 7.4, 0.15 M NaCl, 3.4 mM EDTA, and 0.05% Surfactant P20 (Pharmacia Biosensor)].

Gene Assembly

The capturing of the 5'-biotinylated oligonucleotide 1 to the streptavidin surface was performed in HBS with 0.3 M NaCl, by an injection of 30 μl (2 pmol/μl). Subsequently, 30 μl of oligonucleotide 2 was injected at 2 and 4 pmol/μl. Oligonucleotides 3/4 (4 was nonphosphorylated) and 5/6 were pairwise prehybridized to create the double-stranded middle and outer fragments used for ligation. Each oligonucleotide was first heated to 70°C and then mixed stoichiometrically with the complementary oligonucleotide followed by slow cooling to room temperature. For ligations, 40 μl of the fragments was injected at concentrations of 2 pmol/μl, respectively, in One Phor-All-Buffer Plus, Pharmacia Biotech (10 mM Tris–acetate, pH 7.5, 10 mM magnesium acetate, 50 mM potassium acetate) supplemented with ATP and T4 DNA ligase (Pharmacia Biotech) to final concentrations of 1 mM and 0.1 Weiss U/μl, respectively. After each ligation step, 8 μl of 0.05% sodium dodecyl sulfate (SDS) was injected to remove accumulated ligase from the surface. For strand-specific elution of the nonbiotinylated strand in the assembled DNA fragment, 8 μl of 50 mM sodium hydroxide was used. As control, an injection of a noncomplementary fragment (oligonucleotides 5/6) together with T4 DNA ligase over a surface containing the immobilized 1/2 oligonucleotide pair was performed.

Hybridization Kinetics

In a series of subsequent experiments, three 20-μl pulses, respectively, of oligonucleotides 4 and 6 (0.5, 1, and 2 pmol/μl) were injected over a sensor surface with 810 RU of 69-nt single-stranded DNA obtained from a gene assembly experiment. The hybridization template was regenerated for each cycle by injection of an 8-μl pulse of 50 mM sodium hydroxide.

DNA Synthesis

DNA polymerase I (Klenow fragment, Boehringer–Mannheim) or T7 DNA polymerase (Pharmacia Biotech) was injected (10 μl) at concentrations of 0.05 U/μl in an extension buffer (28 mM Tris–HCl, pH 7.2, 30 mM citric acid, 10 mM $MgCl_2$, 32 mM dithiothreitol, 4 mM $MnCl_2$), supplemented with all four dNTPs to final concentrations of 0.2 mM. The substrates for the extensions were prepared by injections of 20 μl (2 pmol/μl) of the outer oligonucleotide 6 over a sensor surface with 810 RU of 69-nt single-stranded DNA, obtained from a gene assembly experiment. These injections resulted in the hybridization of oligonucleotide 6, which served as primer for the extensions. The single-stranded DNA was regenerated between the runs with 8 μl of 50 mM sodium hydroxide, to elute extended primers. Injections of SDS (0.05%) were used to dissociate the enzyme from surface-bound DNA, prior to the strand separation with alkali.

Cleavage with Endonuclease

A substrate for endonuclease cleavage on the sensor surface was prepared by first injecting 20 μl of oligonucleotide 6 (2 pmol/μl) over approximately 760 RU of full-length (69 nt) single-stranded DNA, obtained from a gene assembly experiment. This resulted in 250 RU of annealed primer. After extension, using Klenow polymerase and dNTPs, the signal increased by 430 RU, resulting in a calculated amount of 1310 RU of double-stranded DNA to serve as endonuclease substrate. A 45-μl injection (1 μl/min) of the endonuclease *Xho*I (Pharmacia Biotech) at a concentration of 0.6 U/μl in 10 mM Tris–HCl, pH 7.5, 0.15 M NaCl, 10 mM $MgCl_2$, and 0.1 μg/μl bovine serum albumin was made over the sensor surface for cleavage at the recognition sequence positioned 16–21 nucleotides from the biotinylated end of the double-stranded DNA fragment. After cleavage, 8 μl of 0.05% SDS was injected to remove eventually accumulated enzyme from the surface.

DNA Minisequencing

A single-stranded template for DNA minisequencing was prepared by a first assembly of the 69-bp model DNA, initiated by the capture of 5'-biotinylated oligonucleotide 6 on a prewashed streptavidin sensor chip (SA5, Pharmacia Biosensor) (data not shown). For elution of the nonbiotinylated complementary DNA strands and regeneration of the templates between each cycle of sequencing, an injection of 8 μl of 50 mM sodium hydroxide was performed. Each cycle was initiated by an injection (20 μl) of nonbiotinylated oligonucleotide 1 [0.5 pmol/μl] for hybridization and use as primers for the extension. In each of the four runs, a 10-μl pulse of 0.05 U/μl Klenow DNA polymerase together with one of the four dideoxynucleotides ddATP/ddTTP/ddCTP/ddGTP (1 μM) in extension buffer (see above) was injected for eventual incorporation and chain termination. SDS (8 μl, 0.05%) was injected after each enzymatic step. The extensions were performed as described above.

RESULTS

The Model System

To analyze different molecular biology techniques in a quantitative and kinetic manner using real-time BIA, a model system comprising six oligonucleotides was designed. These oligonucleotides outlined in Fig. 1 can be assembled into a 69-bp double-stranded DNA fragment

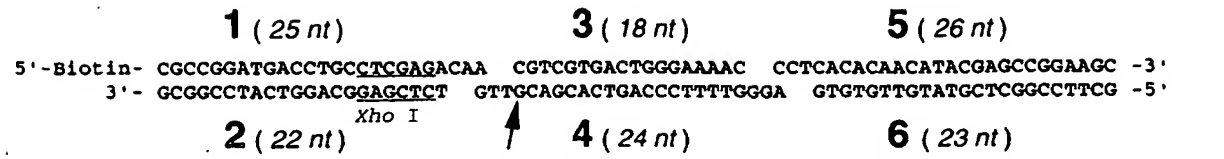


FIG. 1. Sequences, lengths, and relations of the synthetic oligonucleotides. For simplicity the oligonucleotides are denoted 1 through 6. The three subfragments can be assembled into a 69-bp double-stranded DNA fragment via the 3-nt protrusion ends. The G in oligonucleotide 4 marked with an arrow indicates the nucleotide determined in the minisequencing protocol, in which a 5'-biotinylated version of oligonucleotide 6 was used.

from three smaller fragments with overlapping protrusions. The experiments were monitored in real-time using biosensor technology adapted for streptavidin–biotin chemistry. The oligonucleotides were used in different combinations in a number of separate experiments which are described below.

Gene Assembly

As outlined in Figs. 2A–2E, a series of experiments was performed to monitor the assembly of a 69-bp DNA fragment in a stepwise manner using the sensor chip surface as a solid-phase anchor. To obtain an efficient and robust capture of the first DNA segment onto the sensor chip, the well-documented streptavidin–biotin interaction with a $K_d = 10^{-15}$ M (9) was employed. A prewashed sensor chip SA5 containing approximately 3500 RU of covalently coupled streptavidin was used to capture 5'-biotinylated oligonucleotides (Fig. 2A), injected over the surface. The result shows that the capture of this oligonucleotide (Fig. 3A) is rapid and a steady level is quickly reached. The amount bound (1400 RU) corresponds to a stoichiometry of 2.9 bound oligonucleotides (M_r = 8.3 kDa) per tetramer of streptavidin (M_r = 60 kDa, for the tetramer), which theoretically can bind four biotin molecules. This shows that a good accessibility to the binding sites of streptavidin is achieved in the immobilization. Injections with a twofold higher concentration of the oligonucleotide did not increase the uptake (data not shown), indicating a saturation of the streptavidin surface. The captured single-stranded biotinylated 25-mers were subsequently used for the annealing of the complementary strand (Fig. 2B). Oligonucleotide 2 was injected and the uptake was registered. As can be seen in the first part of the injection in the resulting sensorgram (Fig. 3B), the binding of this interaction is slower than that for the streptavidin–biotin interaction. The increase in response (760 RU) corresponds to an apparent efficiency in this step of 64%. Injections with higher concentrations of the oligonucleotide did not result in higher sig-

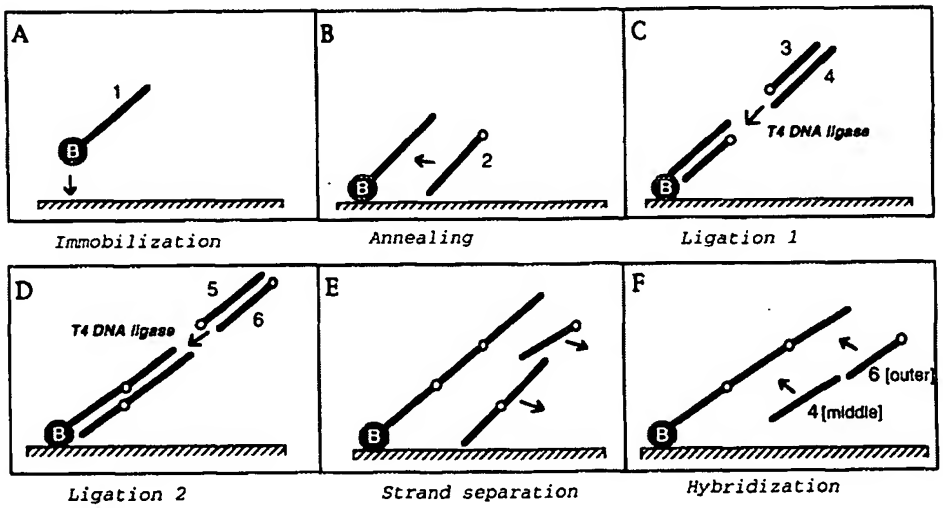


FIG. 2. Schematic view of the consecutive steps in the gene assembly. F shows the subsequent reannealing of oligonucleotides 4 and 6 for the hybridization kinetics study (see text for details). 5'-Phosphate groups of the oligonucleotides are indicated (open circle).

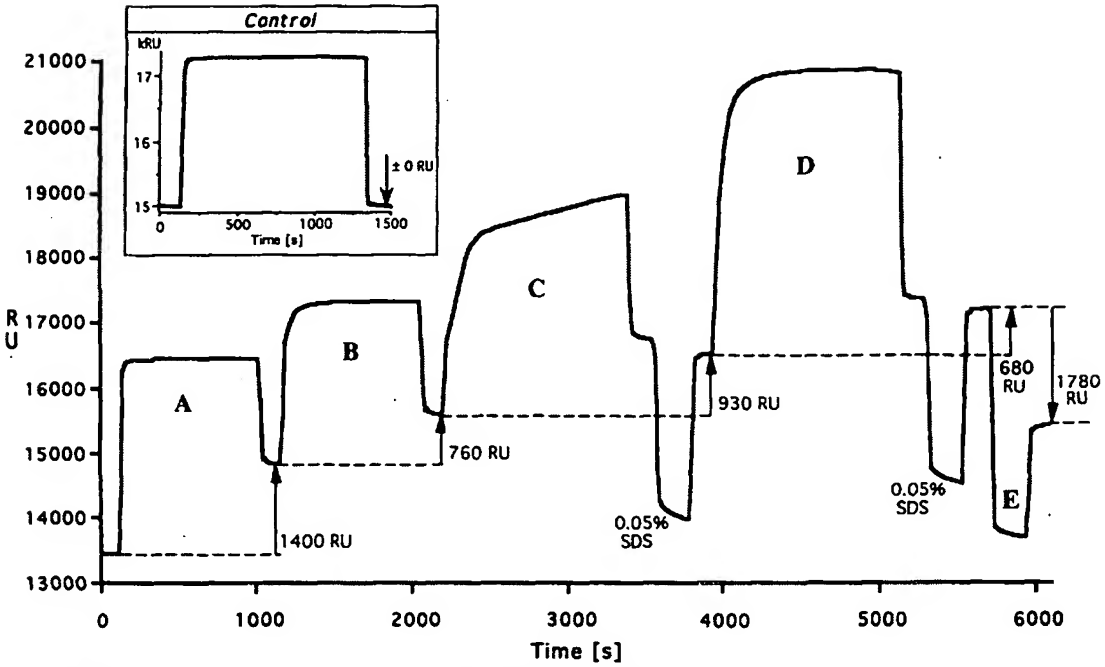


FIG. 3. Sensorgram from the biosensor analysis of the gene assembly. The capital letters correspond to the different events as outlined in the legend to Fig. 2. The numbers in the sensorgram correspond to the net response obtained in each step. Inset: Resulting sensorgram from a control experiment where a noncomplementary double-stranded DNA fragment was injected together with T4 DNA ligase.

nals (data not shown). No detectable binding was seen from injections with a noncomplementary oligonucleotide (data not shown), indicating that the observed annealing was specific.

To continue the assembly of the DNA fragment, oligonucleotides 4 (nonphosphorylated) and 3 were prehybridized and subsequently injected over the surface. This fragment contains a 3-nt protrusion which is complementary to the sequence present at the free end of the immobilized fragment, consisting of oligonucleotides 1 and 2. No increase in the signal was observed when the injection was performed without DNA ligase (data not shown). In contrast, when injecting the fragment with the addition of T4 DNA ligase (Fig. 2C), a significant binding was recorded by the real-time analysis (Fig. 3C). This shows that the binding via the overlapping region of 3 bp is not sufficient for the formation of stable noncovalent complexes at this temperature (25°C), but the DNA ligase allows a transient interaction followed by covalent attachment. The shape of the curve observed during the injection of the DNA/ligase mixture indicates a biphasic reaction (Fig. 3C). The gradual increase of the signal, at the end of the injection pulse, also suggests that the equilibrium is not reached during the injection period (20 min). A subsequent injection of SDS washing solution resulted in a 200-RU drop in response (Fig. 3, C and D), demonstrating the necessity to eliminate background response from T4 DNA ligase attached to the DNA for correct evaluation of the result. A single injection of the SDS solution was sufficient, since repeated injections did not further affect the signal (data not shown). From the analysis, a net response of 930 RU was calculated, corresponding to a ligation efficiency of 64% (930/1450). In a control experiment, in which a fragment containing a noncomplementary protrusion was injected together with T4 DNA ligase, no change in the signal was observed (Fig. 3, inset).

A second ligation step, completing the assembly, was performed using prehybridized oligonucleotides 5 and 6 (Fig. 2D). Note that oligonucleotide 4 lacks a 5′-phosphate group and ligation can therefore only occur on one of the strands. A plateau value was reached after approximately 13 min (Fig. 3D), suggesting that this ligation step reaches equilibrium more rapidly than the previous ligation (Fig. 3C). Interestingly, the shape of the curve indicates that this ligation is not a biphasic event. This might be explained by the absence of a 5′-phosphate group in oligonucleotide 4, resulting in the formation of a single phosphodiester bond per fragment, rather than two in the preceding ligation. The efficiency of this second ligation step, calculated from the response after the SDS washing (680 RU) was 63%. The assembly was fol-

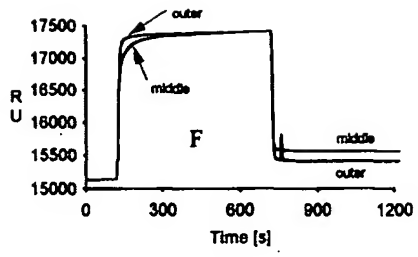


FIG. 4. An overlay plot of sensorgrams from the hybridization experiment. Oligonucleotides 4 and 6 were hybridized onto an immobilized single-stranded DNA template as outlined in the legend to Fig. 2F (see text for details).

lowed by strand-specific elution of the nonbiotinylated strand (Fig. 2E). A short pulse of alkali resulted in a significant decrease in the signal (Fig. 3E). The amount released (1780 RU) suggests efficient release of the nonimmobilized strand. The amount of full-length single-stranded DNA present on the surface after this alkali pulse is approximately 970 RU. Note that the eluted DNA consist of two separate fragments, due to the absence of a phosphate group in oligonucleotide 4.

Hybridization Kinetics

To analyze the influence of the relative distance from the streptavidin/dextran surface on the hybridization kinetics, the hybridization of two oligonucleotides complementary to different regions of an immobilized DNA strand was studied (Fig. 2F). Equimolar amounts of oligonucleotides 4 and 6, respectively, were injected separately over immobilized single-stranded DNA obtained from a gene assembly experiment. An overlay plot of the resulting sensorgrams shows a clear difference in the hybridization rate of the two oligonucleotides (Fig. 4). Interestingly, oligonucleotide 6, hybridizing to the sequence localized in the 3'-end (outermost sequence) of the immobilized single-stranded DNA, shows the fastest binding kinetics. The hybridization of the corresponding oligonucleotide to the middle sequence, oligonucleotide 4, shows a slower rate. This indicates that there are factors (i.e., sterical or electrostatic) influencing the hybridization, which might correlate to the relative position of annealing on an immobilized single-stranded DNA fragment. The dissociation (off-rate) of the hybridized oligonucleotides from the immobilized DNA is slow for both oligonucleotides (Fig. 4), regardless of the relative annealing position. Therefore, the difference in annealing kinetics is related only to the on-rate of the interaction, since the rate of annealing is dependent on both the on- and off-rates. The observed amounts of hybridized oligonucleotides corresponded to an apparent saturation of the template at the different positions, since injections with lower concentrations resulted in the same steady-state values (data not shown). For the outermost position (oligonucleotide 6), a maximum hybridization of 98% of the theoretical value was achieved, whereas for the middle position (oligonucleotide 4), the value was lower (84%). This further shows a dependency in the hybridization of factors related to the relative distance from the anchoring biotin.

DNA Synthesis

The results from the gene assembly experiment show that an enzymatic action upon DNA can be monitored with real-time BIA, using standard flow rates and injection volumes. To investigate if the performance of other nucleic acid-modifying enzymes could also be monitored using BIA, the polymerase activities of T7 DNA polymerase and DNA polymerase I (Klenow fragment) were investigated. To prepare suitable substrates for both of these primer-dependent enzymes oligonucleotide 6 was hybridized to the 3'-end of the immobilized 69-nt single-stranded DNA (Fig. 5A-I). After injection of T7 DNA polymerase in a buffer containing dNTPs (Fig. 5A-II), a gradual increase of the signal was seen during the entire sample pulse. In the following flow of running buffer, a slow dissociation from the surface was observed (Fig. 5B-II). A subsequent pulse of SDS solution resulted in a further decrease to a stable value (Fig. 5B), which was not changed by further injections. In a control experiment, in which T7 DNA polymerase was injected over a plain streptavidin surface, no interaction was observed (data not shown), indicating that the previous interaction was DNA dependent. When T7 DNA polymerase was injected over the single-stranded DNA/primer substrate, without nucleotides present in the buffer, a gradual accumulation was seen, but with no increase of the signal after SDS washing (data not shown). Therefore, the gradual increase of the signal seen during the entire pulse of T7 DNA polymerase in the extension experiment can be interpreted as a sum of signals from extension and accumulation of T7 DNA polymerase on the DNA. The resulting SPR response corresponds to a near complete extension of all the hybridized primers. A final injection of alkali results in a decrease of approximately 680 RU, which correlates well with a complete release of the "second-strand" DNA made up from extended primers (Fig. 5B-III).

In a subsequent study, DNA polymerase I (Klenow fragment) was analyzed using the same experimental setup. For this polymerase, a gradual increase of the RU signal was seen only during the first minute of injection, after which the signal rapidly declines to a steady plateau value (Fig. 5C-II). In the following flow of running buffer, no significant decline was seen and the subsequent SDS pulse only resulted in a minor decrease of the

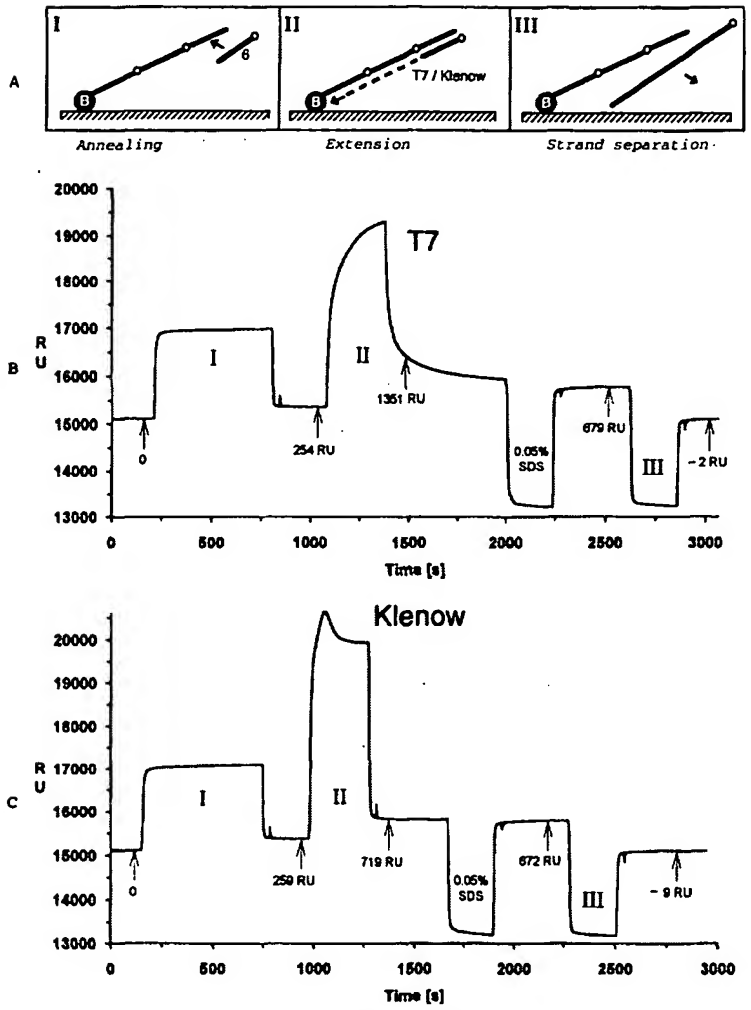


FIG. 5. Sensorgram from the DNA synthesis experiment. The DNA polymerase activities of T7 DNA polymerase and *Escherichia coli* DNA polymerase I (Klenow fragment) were investigated in real-time in a primer extension experiment. (A) Schematic overview of the steps in the experiment; (B) sensorgram obtained from the experiment using T7 DNA polymerase; (C) sensorgram obtained from the experiment using DNA polymerase I (Klenow fragment); see text for details.

signal (Fig. 5C). Also in this case, the resulting signal decrease after the final alkali injection corresponds to a complete release of the second-strand DNA, obtained from an extension of the hybridized primers (approximately 680 RU). The complete dissociation from the immobilized templates seen for the Klenow polymerase already after approximately 1 min of injection indicates that the DNA synthesis is completed after that period of time.

Endonuclease Cleavage

Oligonucleotide 1 contains the contribution from one strand to the recognition sequence for the endonuclease *Xho*I (Fig. 1). Thus, double-stranded DNA obtained by extension toward the sensor chip surface using the priming outer oligonucleotide 6 together with Klenow polymerase contains the complete recognition site (Fig. 6A). To evaluate if an endonuclease activity could also be monitored using BIA, a 45-min pulse of *Xho*I was injected over 1310 RU of such 69-bp substrate DNA (Fig. 6B), obtained from a gene assembly experiment (data not shown). This resulted in an initial increase of the signal, followed by a slow but significant decline. After the injection of endonuclease, the response level decreased with approximately 900 RU. A subsequent SDS pulse did not result in any further decrease. Taking into

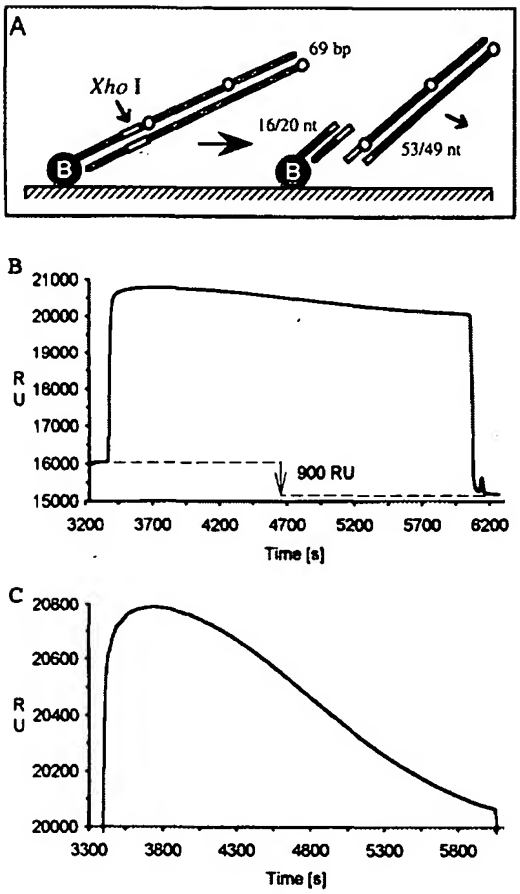


FIG. 6. Cleavage with endonuclease. (A) A schematic view of the cleavage of the immobilized double-stranded DNA fragment; (B) sensorgram registered during the injection of the endonuclease *Xho* I; (C) enlargement of the upper section of the sensorgram in Fig. 6B (see text for details).

account that a fragment of 16/20 nt is left on the surface after cleavage, the efficiency at this temperature (25°C) was calculated at 93%. An enlargement of the sensorgram corresponding to the cleavage phase clearly demonstrates a sigmoidal curve shape for this enzyme–substrate reaction (Fig. 6C).

DNA Minisequencing

The results from the extension experiments implied the possibilities to develop a strategy to employ the resulting SPR signal for a qualitative analysis of the template DNA strand. An incorporation of one of four dideoxynucleotides (ddNTPs) at the position adjacent to the primer in a first step would prevent any further extension in a subsequent step using all four dNTPs. Thus, the identity of the base pair immediately downstream of the primer would be revealed by the absence of a signal related to extension in the second step. This concept of real-time DNA minisequencing was tested using single-stranded full-length 69-bp DNA fragments obtained from a gene assembly followed by alkaline elution. For this experiment, the assembly was performed using a biotinylated version of the oligonucleotide 6 as starting fragment, thus resulting in DNA fragment as outlined in Fig. 7A. Using the same sensor chip surface, four consecutive sensorgrams were registered (Fig. 7B), each representing the successive injections of (i) oligonucleotide 1; (ii) Klenow polymerase and one of the four ddNTPs; (iii) SDS; (iv) Klenow polymerase and all four dNTPs; (v) SDS; and (vi) alkali. A small gradual decrease in the amount of hybridized oligonucleotide 1 can be seen for the consecutive runs, which can be explained by a corresponding small loss of template during the alkaline pulses. A comparison of the extension phases following the injections of Klenow polymerase and one ddNTP reveals that three of the four curves (corresponding to injections with ddGTP, ddTTP, and ddATP) contain the characteristic "extension/dissociation" peak observed earlier for extension employing Klenow polymerase (see DNA Synthesis). In addition, the resulting signals after SDS pulses show increases in the signal corresponding to an extension of the hybridized primers. Furthermore, the subsequent alkali pulses results in an elution of the extended strands. Interestingly, in the case where the elongation terminator ddCTP was injected together with Klenow polymerase, the curve shape in the following extension step is dramatically different. No characteristic peak could be observed and a significantly slower dissociation from the surface was observed. After the subsequent SDS pulse only a small increase in the signal was observed. This indicates that ddCTP was incorporated by the polymerase at the position adjacent to the primer (oligonucleotide 1), thus preventing extension in the following step. The results from the experiment using ddCTP nucleotides suggest that the corresponding position on the complementary strand is a G, which could be verified from the sequences of the oligonucleotides used (Fig. 1).

DISCUSSION

In this work we have shown the possibilities of using biosensor technology for quantitative and qualitative analysis in real-time of several molecular biology techniques. The streptavidin–biotin system was used for efficient capture and robust immobilization of biotinylated nucleic acids onto the sensor chip surface. For small fragments (e.g., 25-mers) used in this study, a rapid and high-level immobilization can routinely be obtained, using standard buffers and flow rates. A clear

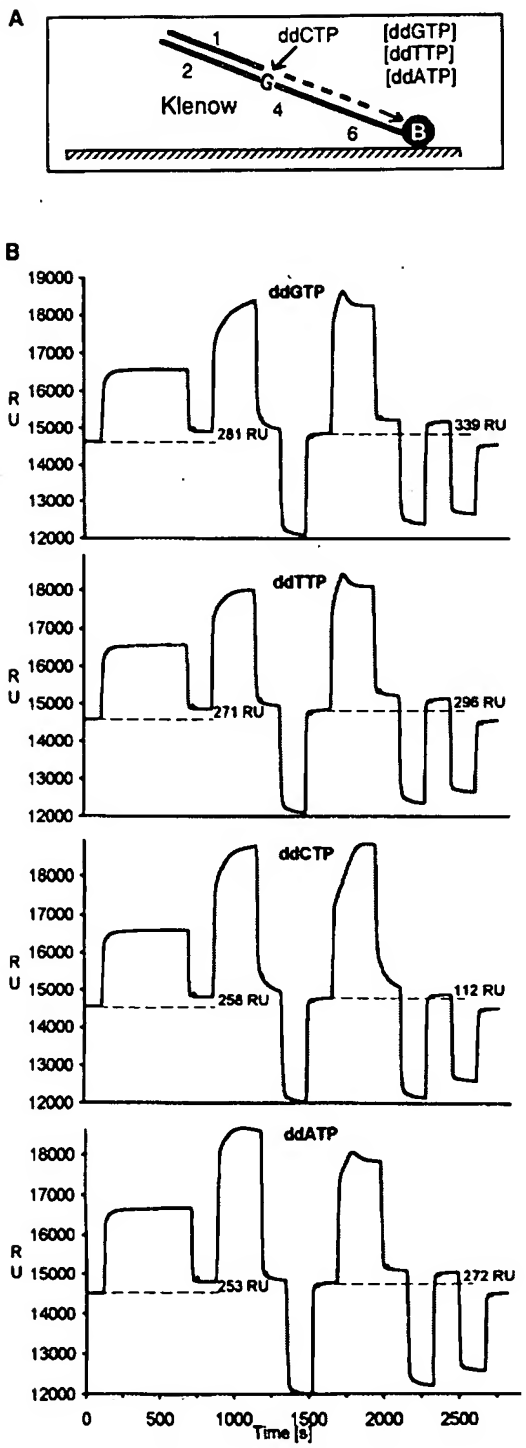


FIG. 7. DNA minisequencing. (A) A schematic description of the principle. (B) Sensorgrams from the four consecutive runs with injections of (i) oligonucleotide 1 for hybridization; (ii) one of the four dideoxynucleotides (indicated) together with Klenow DNA polymerase; (iii) 0.05% SDS; (iv) Klenow DNA polymerase with all four dNTPs; (v) 0.05% SDS; and (vi) alkali for strand separation and regeneration of the single-stranded DNA template.

difference in the curve shapes was observed for the two ligation steps in the gene assembly experiment. These ligation steps take place at different relative distances from the surface and involve the formation of two and one phosphodiester bond(s), respectively. However, when analyzing results from similar experiments (data not shown), a general two-phasic curve for ligations involving the formation of two phosphodiester bonds can be seen, regardless of the position for ligation. The successful real-time monitoring of the assembly of the 69-bp double-stranded model DNA fragment shows the possibilities of using biosensor technology for the optimization of critical steps during the procedure, which is not possible for other solid-phase gene assembly methods (10–12). The short protrusion ends (3 nt) of the fragments were sufficient for the formation of substrates for the ligase, even at 25°C, a temperature well above the calculated melting temperatures (T_m) for the annealing stretches (7 and 8°C, respectively). After completed assembly, the full-length DNA can be released from the surface by either alkali or restriction with endonucleases for collection at the outlet of the flow system and eventual PCR amplification prior to cloning. The technique opens the possibilities to monitor and optimize critical steps in the assembly of overlapping hexameric oligonucleotides into lengths feasible for use as sequencing primers for large-scale sequencing projects employing "primer-walking" strategies (13).

The analysis of the DNA polymerase activities of Klenow DNA polymerase and T7 DNA polymerase suggests that the technique can be used for the characterization of different polymerases according to several features such as extension rate, influence of DNA context, promiscuity for nucleotide analogues (e.g., azidothymidine, ddI, alpha- and deaza-nucleotides), temperature dependency, and the effects of inhibitors [e.g., anti-*Taq* polymerase antibodies (14) for hot-start PCR]. The data from the *Xho*I endonuclease cleavage experiment show also that for this class of enzymes, several important features can potentially be evaluated in real-time, using an experimental setup employing biosensor technology.

The DNA minisequencing experiment shows that several different unit operations easily can be functionally integrated. The background extension signal (112 RU) obtained in the cycle where the dideoxynucleotide was incorporated (ddCTP) could be explained by the 3′-5′ exonuclease activity of Klenow polymerase (proofreading). This activity results in the removal of a fraction of the incorporated dideoxynucleotides, leading to a free 3′-end

as substrate for the preceding extension phase. Thus, the use of a DNA polymerase devoid of this activity (15) might prevent this background signal. The slow dissociation from the surface seen for the Klenow polymerase in this experiment (Fig. 7B-ddCTP) is not fully understood, but a correlation to the proofreading activity cannot be excluded. The minisequencing strategy described here has the advantages to be fully automated with no need for labeling. This method can potentially be applied also for multipoint determinations on single-stranded DNA templates, originating from PCR amplification of, e.g., clinical samples. Here, sequential hybridizations and extensions could be performed using different priming oligonucleotides for each of the selected positions. In conclusion, these results suggest that real-time interaction analysis using biosensor technology is a powerful tool to characterize enzymes and reactions used for nucleic acid manipulation. Several future applications in the field of sequencing by synthesis or sequencing by hybridization using PCR products can also be envisioned.

ACKNOWLEDGMENTS

This work was supported by grants from the Swedish Natural Science Research Council. Drs. J. Odeberg and J. Lundeberg are acknowledged for valuable discussions and advice and A. Larsson and J. Richey for critical reading of the manuscript.

REFERENCES

1. Jönsson, U., Fägerstam, L., Ivarsson, B., Johnsson, B., Karlsson, R., Lundh, K., Löfås, S., Persson, B., Roos, H., Rönnberg, I., Sjölander, S., Stenberg, E., Ståhlberg, R., Urbaniczky, C., Östlin, H., and Malmqvist, M. (1991) *BioTechniques* **5**, 620–627.
2. Löfås, S., Malmqvist, M., Rönnberg, I., Stenberg, E., Liedberg, B., and Lundström, I. (1991) *Sensors Actuators B* **5**, 79–84.
3. Karlsson, R., Fägerstam, L., Nilshans, H., and Persson, B. (1993) *J. Immunol. Methods* **166**, 75–84.
4. Bondeson, K., Frostell-Karlsson, Å., Fägerstam, L., and Magnusson, G. (1993) *Anal. Biochem.* **214**, 245–251.
5. Fisher, R. J., Fivash, M., Casas-Finet, J., Erickson, J. W., Kondoh, A., Bladen, S. V., Fisher, C., Watson, D. K., and Papas, T. (1994) *Protein Sci.* **3**, 257–266.
6. Wood, S. J. (1993) *Microchem. J.* **47**, 330–337.
7. Hultman, T., Berg, S., Moks, T., and Uhlén, M. (1991) *BioTechniques* **1**, 84–93.
8. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, second ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
9. Green, N. M. (1975) *in* Advances in Protein Chemistry (Anson, M. L., and Edsell, J. T., Eds.), Vol. 29, pp. 85–133, Academic Press, NY.
10. Hostomsky, Z., Smrt, J., Arnold, L., Tocik, Z., and Paces, V. (1987) *Nucleic Acids Res.* **12**, 4849–4856.
11. Ståhl, S., Hansson, M., Ahlborg, N., Nguyen, T. N., Liljeqvist, S., Lundeberg, J., and Uhlén, M. (1993) *BioTechniques* **3**, 424–434.
12. Nguyen, T. N., Uhlén, M., and Ståhl, S. (1994) *in* Advances in Biomagnetic Separation (Uhlén, M., Hornes, E., and Olsvik, Ø., Eds.), pp. 73–78, Eaton, IL.
13. Strauss, E. C., Kobori, J. A., Siu, G., and Hood, L. E. (1986) *Anal. Biochem.* **154**, 553–560.
14. Sharkey, D. J., Scalice, E. R., Christy, K. G., Jr., Atwood, S. M., and Daiss, J. L. (1994) *Bio/Technology* **12**, 506–509.
15. Tabor, S., and Richardson, C. C. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 4767–4771.

US005795714A

United States Patent [19]

Cantor et al.

[11] Patent Number: 5,795,714

[45] Date of Patent: Aug. 18, 1998

[54] **METHOD FOR REPLICATING AN ARRAY OF NUCLEIC ACID PROBES**

[75] Inventors: **Charles R. Cantor; Marek Przetakiewicz; Cassandra L. Smith; Takeshi Sano**, all of Boston, Mass.

[73] Assignee: **Trustees of Boston University**, Boston, Mass.

[21] Appl. No.: **110,691**

[22] Filed: **Aug. 23, 1993**

Related U.S. Application Data

[63] Continuation-in-part of Ser. No. 972,012, Nov. 6, 1992, Pat. No. 5,503,980.

[51] **Int. Cl.**[6] **C12Q 1/68**

[52] **U.S. Cl.** **435/6**; 435/6

[58] **Field of Search** 435/6, 91.1; 536/24.3, 536/24.33, 25.3

[56] **References Cited**

U.S. PATENT DOCUMENTS

4,808,520	2/1989	Dattagupta et al. .	
5,002,867	3/1991	Macevicz	435/6
5,068,176	11/1991	Vijg et al.	435/6
5,073,483	12/1991	Lebacq	435/6
5,106,727	4/1992	Hartley et al.	435/6
5,112,734	5/1992	Kramer et al.	435/6
5,112,736	5/1992	Caldwell et al.	435/6
5,114,839	5/1992	Blocker	435/6
5,137,806	8/1992	LaMaistre et al. .	
5,149,625	9/1992	Church et al.	435/6
5,202,231	4/1993	Drmanac et al.	435/6
5,219,726	6/1993	Evans	435/6

FOREIGN PATENT DOCUMENTS

0392546	12/1990	European Pat. Off. .
WO89 10977	11/1989	WIPO .

OTHER PUBLICATIONS

Hames, "Nucleic acid hybridisation" (1985), pp. 17–45.

Matthews, et al., "Analytical Strategies for the Use of DNA Probes", (1988) Anal. Biochem 169, pp. 1–25.

STRATAGENE Catalog, "Synthetic Oligonucleotides" (1992), p. 106.

Article, "Optical Properties of Specific Complexes Between Complementary Oligoribonucleotides", Robert B. Gennis et al. *Biochemistry*, vol. 9, No. 24, 1970.

Article, "DNA Sequencing By Hybridization: 100 Bases Read By a Non-gel-based Method", Zaklina Strezoska, *Proc. Natl. Acad. Sci. USA*, vol. 88, pp. 10089–10093, Nov. 1991.

Article, "Improved Chips for Sequencing by Hybridization", P.A. Pevzner, et al., *Journal of Biomolecular Structure & Dynamics*, vol. 9, Issue No. 2, 1991.

Article, "A Method for DNA Sequencing By Hybridization with Oligonucleotide Matrix", K.R. Khrapko et al. *DNA Sequence –J. DNA Sequencing and Mapping*, vol. 1, pp. 375–388, 1991.

Article, "Sequence-Selective Recognition of DNA by Strand Displacement with a Thymine-Substituted Polyamide", Peter E. Neilsen et al. *Science*, vol. 254, Dec. 1991.

Article, Parallel Analysis of Oligodeoxyribonucleotide (oligonucleotide) interactions. I. Analysis of Factors Influencing Oligonucleotide Duplex Formation, Uwe Maskos et al. *Nucleic Acids Research*, vol. 20, No. 7, pp. 1675–1678, 1992.

(List continued on next page.)

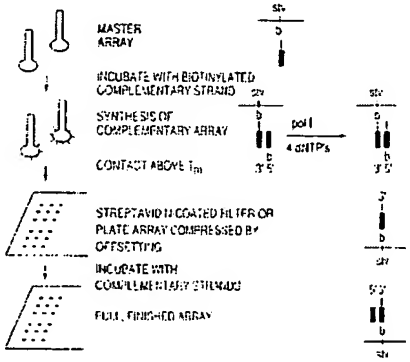
Primary Examiner—Scott W. Houtteman

Attorney, Agent, or Firm—James Remenick; Baker & Botts, L.L.P.

[57] **ABSTRACT**

The invention relates to the replication of probe arrays and methods for replicating arrays of probes which are useful for the large scale manufacture of diagnostic aids used to screen biological samples for specific target sequences. Arrays created using PCR technology may comprise probes with 5'- and/or 3'-overhangs.

29 Claims, 13 Drawing Sheets



OTHER PUBLICATIONS

Article. Oligonucleotide Hybridisations of Glass Supports: A Novel Linker for Oligonucleotide Synthesis and Hybridisation Properties of Oligonucleotides Synthesised in Situ. Uwe Maskos. *Nucleic Acids Research*. vol. 20. No. 7. pp. 1679–1684. 1992.

Article. "Encoded Combinatorial Chemistry". Sydney Brenner. *Proc. Natl. Acad. Sci. USA*. vol. 89. pp. 5381–5383. Jun. 1992.

Article. "Report on the Sequencing by Hybridization Workshop". C.R. Cantor. *Genomics* 13. 1378–1383. 1992.

1-Tuple DNA Sequencing: Computer Analysis. Pavel A. Pevzner et al. *Journal of Biomolecular Structure & Dynamics*. vol. 7. No. 1 (1989). pp. 63–69.

Reliable Hybridization of Oligonucleotides as Short as Six Nucleotides. R. Drmanac et al. *DNA and Cell Boilogy*. vol. 9. No. 7. 1990 pp. 527–534.

"Analytical Strategies for the Use of DNA Probes". Jayne A. Matthews. *Biochemistry*. vol. 169. (1988). pp. 1–25.

"Rapid Detection and Sequencing of Specific in vitro Amplified DNA Sequences Using Solid Phase Methods". Johan Wahlberg et al. *Mol. Cell. Probes*. vol. 4(4). 1990. pp. 285–297.

A New Method for Determining the DNA Nucleotide Sequence by Hybridization with Ologonucleotides. Yu P. Lysov et al. *Doklady Akademii Nauk SSSR*. vol. 303. No. 6. Dec. 1988. 1508–1511.

"An Oligonucleotide Hybridization Approach to DNA Sequencing". K.R. Khrapko et al. *FEBS Letters*. vol. 256. No. 1–2. Oct. 1989. pp. 118–122.

An Algorithm for the DNA Sequence Generation from k-Tuple Word Contents of the Minimal Number of Random Fragments. D. Drmanac et al. *Journal of Biomolecular Structure & Dynamics*. vol. 8. No. 5. (1991) pp. 1085–1102.

FIG. 1

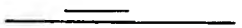
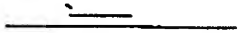
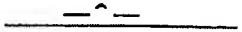
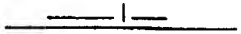
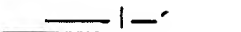
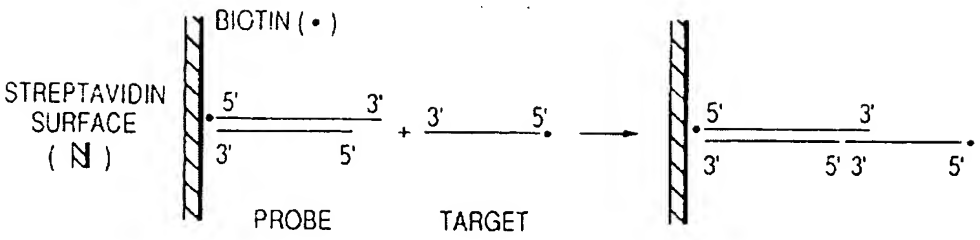
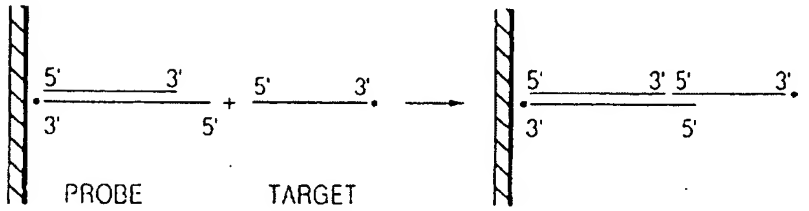
NUCLEIC ACID STRUCTURE	CALCULATED T_m (°C, AVERAGE BASE COMPOSITION)			
	n = 8	7	6	5
	38	33	25	15
	33	25	15	3
	25	15	3	-14
	51	46	40	31
	46	40	31	21
	40	41	21	11

FIG. 2A



BASIC SCHEME FOR POSITIONAL SBH

FIG. 2B



ALTERNATE SCHEME

FIG. 3A

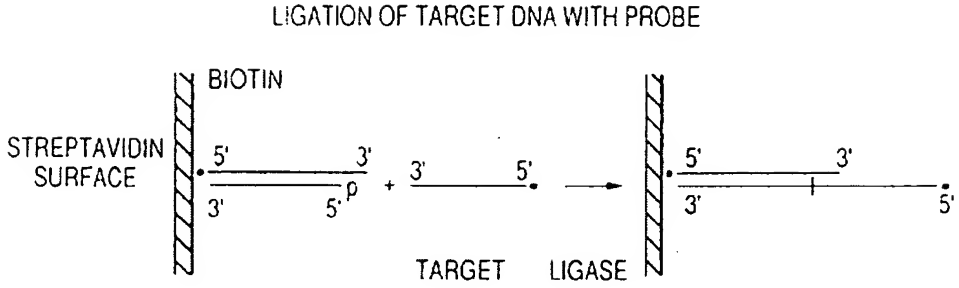

LIGATION OF TARGET DNA WITH PROBE
BIOTIN
STREPTAVIDIN SURFACE
5'
3'
3'
5'
P
3'
5'
TARGET
LIGASE
5'
3'
3'
5'

FIG. 3B

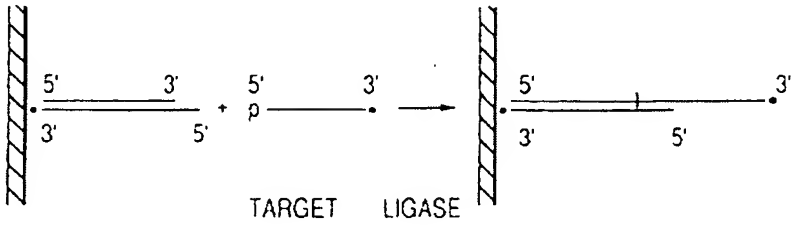

5'
3'
3'
5'
5'
3'
P
TARGET
LIGASE
5'
3'
3'
5'

FIG. 4

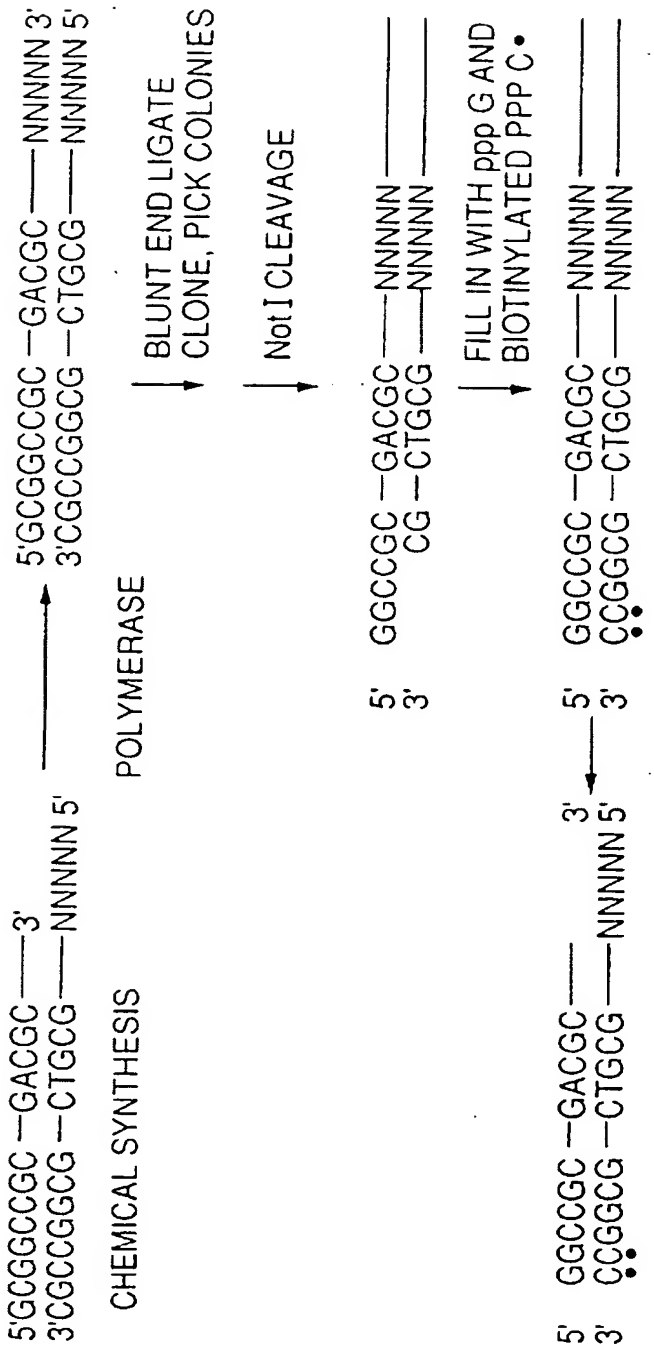
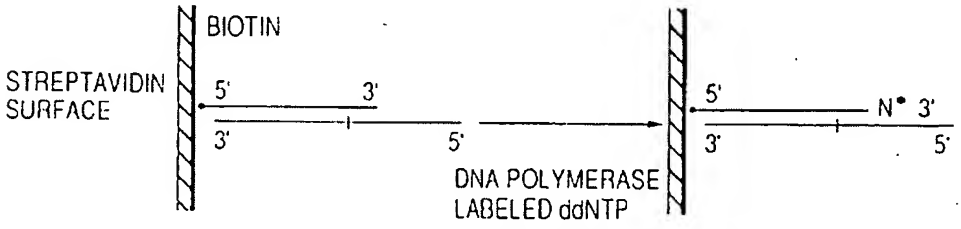
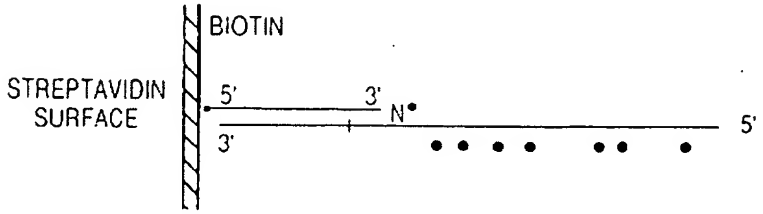

5'GCGGCCGC—GACGC——3'
3'CGCCGGCG—CTGCG——NNNNN 5'
CHEMICAL SYNTHESIS
POLYMERASE
5'GCGGCCGC—GACGC——NNNNN 3'
3'CGCCGGCG—CTGCG——NNNNN 5'
BLUNT END LIGATE
CLONE, PICK COLONIES
Not I CLEAVAGE
5' GGCCGC—GACGC——NNNNN
3' CG—CTGCG——NNNNN
FILL IN WITH ppp G AND
BIOTINYLATED PPP C•
5' GGCCGC—GACGC—— 3'
3' C•CGGCG—CTGCG——NNNNN 5'
CLEAVAGE WITH Hga I
5' GGCCGC—GACGC——NNNNN
3' C•CGGCG—CTGCG——NNNNN

FIG. 5



READING AN EXTRA TARGET BASE BY 3' EXTENSION OF THE PROBE

FIG. 7



POSITIONAL INFORMATION FROM RATIO OF INTERNAL LABEL (•) TO EXTENSION LABEL (•). A 5' LABEL COULD ALSO BE USED.

FIG. 6

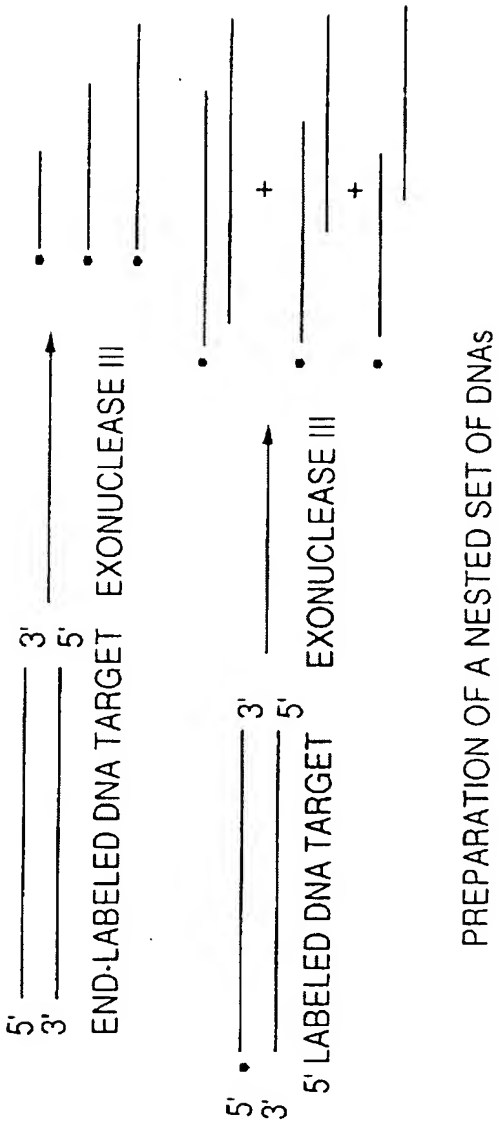

5'
3'
3'
5'
END-LABELED DNA TARGET
EXONUCLEASE III
+
5'
3'
3'
5'
5' LABELED DNA TARGET
EXONUCLEASE III
+
PREPARATION OF A NESTED SET OF DNAs

FIG. 8

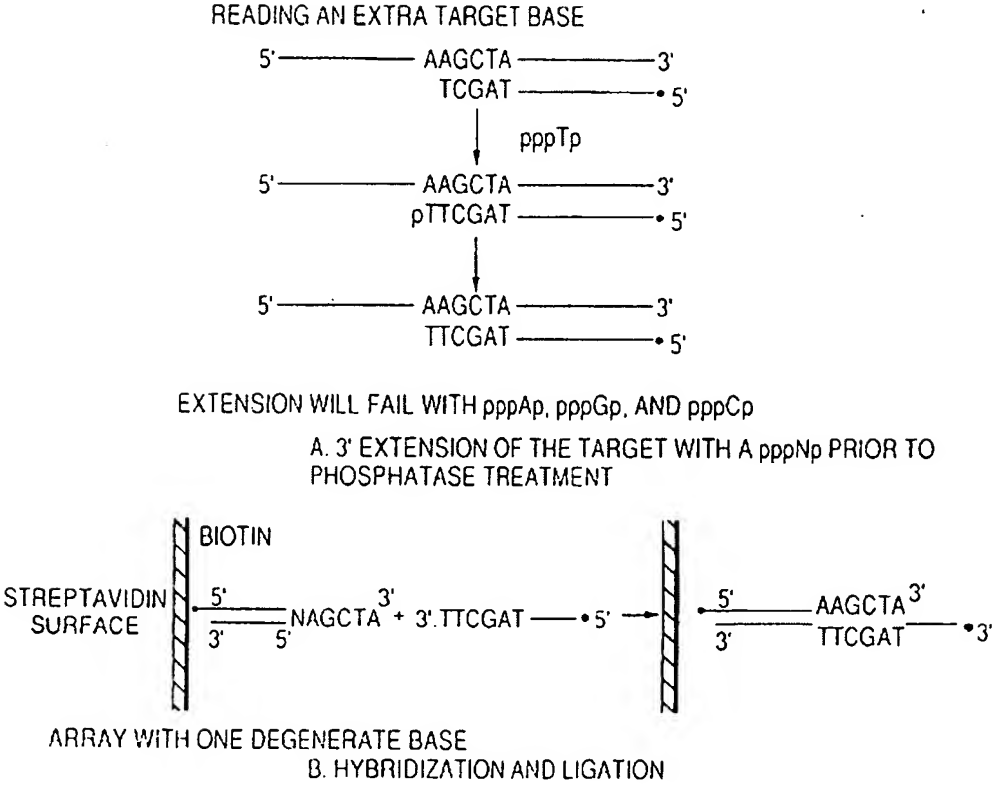

READING AN EXTRA TARGET BASE
5'
AAGCTA
3'
TCGAT
5'
pppTp
5'
AAGCTA
3'
pTTCGAT
5'
5'
AAGCTA
3'
TTCGAT
5'
EXTENSION WILL FAIL WITH pppAp, pppGp, AND pppCp
A. 3' EXTENSION OF THE TARGET WITH A pppNp PRIOR TO PHOSPHATASE TREATMENT
BIOTIN
STREPTAVIDIN SURFACE
5'
3'
NAGCTA
3'
+ 3'.TTCGAT
5'
5'
3'
AAGCTA
3'
TTCGAT
3'
ARRAY WITH ONE DEGENERATE BASE
B. HYBRIDIZATION AND LIGATION

FIG. 9

	LABEL INTENSITY		
CHAIN TERMINATOR:	T	G	A
SEQUENCE			
GGAAT	2A, 2G	–	2G
AAGGT	2A, 2G	2A	–
GAAGT	2A, 2G	–	1G
AGGAT	2A, 2G	1A	–
AGAGT	2A, 2G	1A	–
GAGAT	2A, 2G	–	1G
GGGAT	1A, 3G	–	3G
GGAGT	1A, 3G	–	2G
GAGGT	1A, 3G	–	1G
AGGGT	1A, 3G	1A	–
AAAGT	3A, 1G	3A	–
AAGAT	3A, 1G	2A	–
AGAAT	3A, 1G	1A	–
GAAAT	3A, 1G	–	1G
GGGGT	4G	–	1T, 4G*
AAAAT	4A	1T, 4*	–

FOUR COLOR ANALYSIS OF SEQUENCE EXTENSION OF THE 3' END OF THE PROBE. LABEL INTENSITY SHOWN DOES NOT INCLUDE CONTRIBUTION FROM THE 3' TERMINATOR.

* PLUS ADDITIONAL POSSIBLE RESIDUES

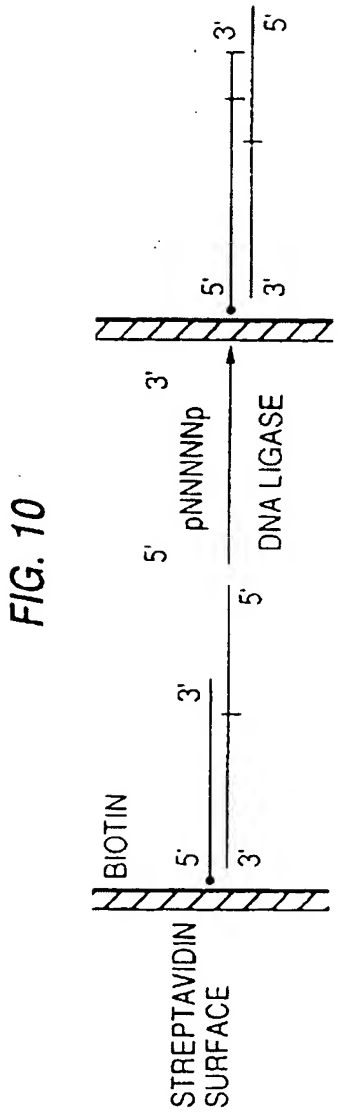
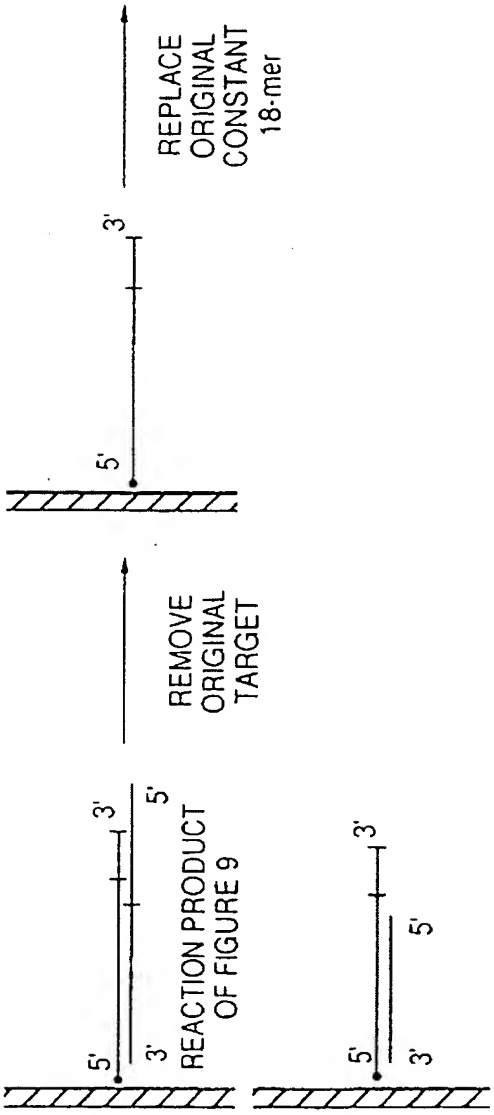


FIG. 10

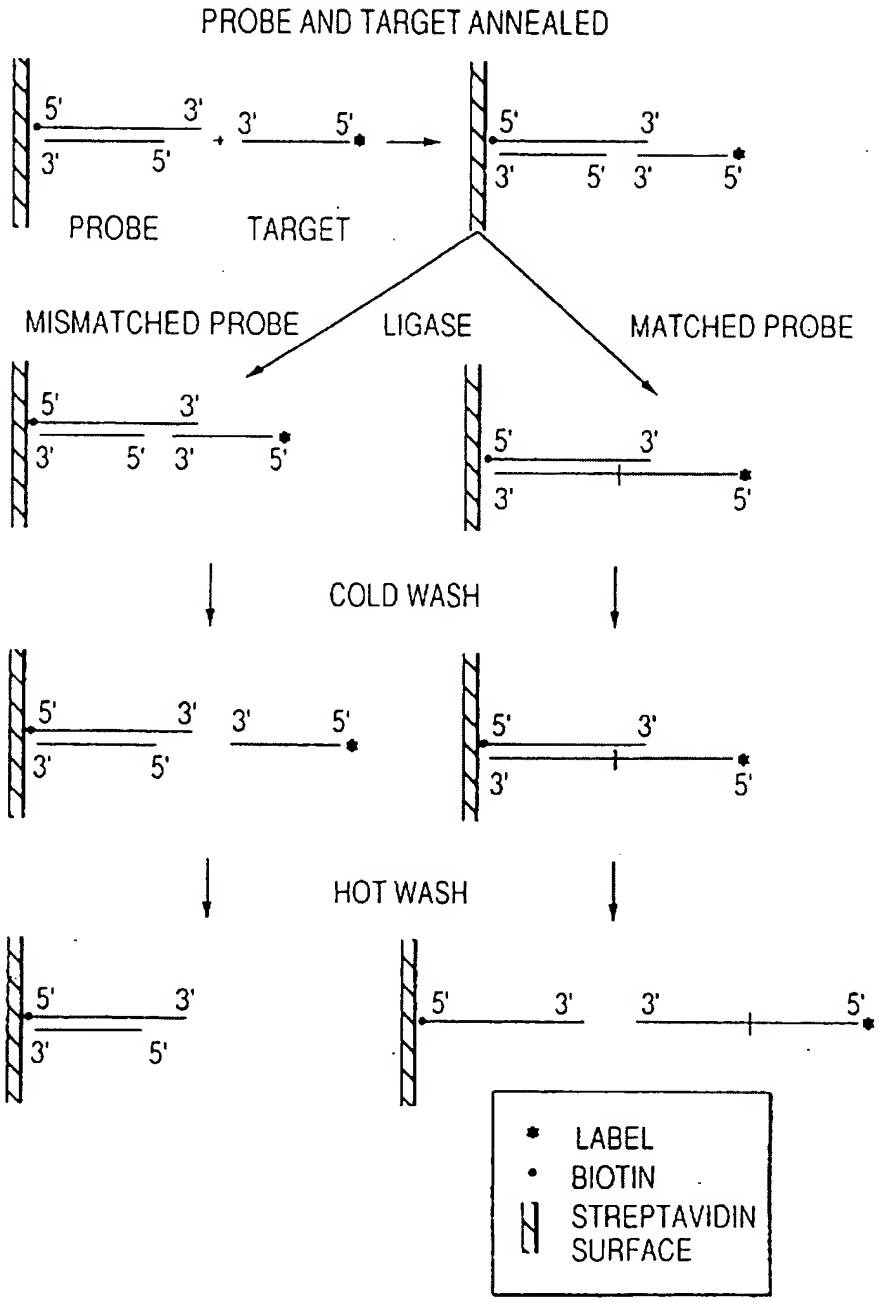
EXTENSION OF THE PROBE BY LIGATION OF A BLOCKED PENTANUCLEOTIDE. THE PENTANUCLEOTIDE IS 3'-BLOCKED TO PREVENT POLYMERIZATION.

FIG. 11



PREPARATION OF A CUSTOMIZED PROBE CONTAINING A 10 bp SEQUENCE PRESENT IN THE ORIGINAL TARGET DNA

FIG. 12


PROBE AND TARGET ANNEALED
5'
3'
3'
5'
PROBE
TARGET
MISMATCHED PROBE
LIGASE
MATCHED PROBE
COLD WASH
HOT WASH
LABEL
BIOTIN
STREPTAVIDIN
SURFACE

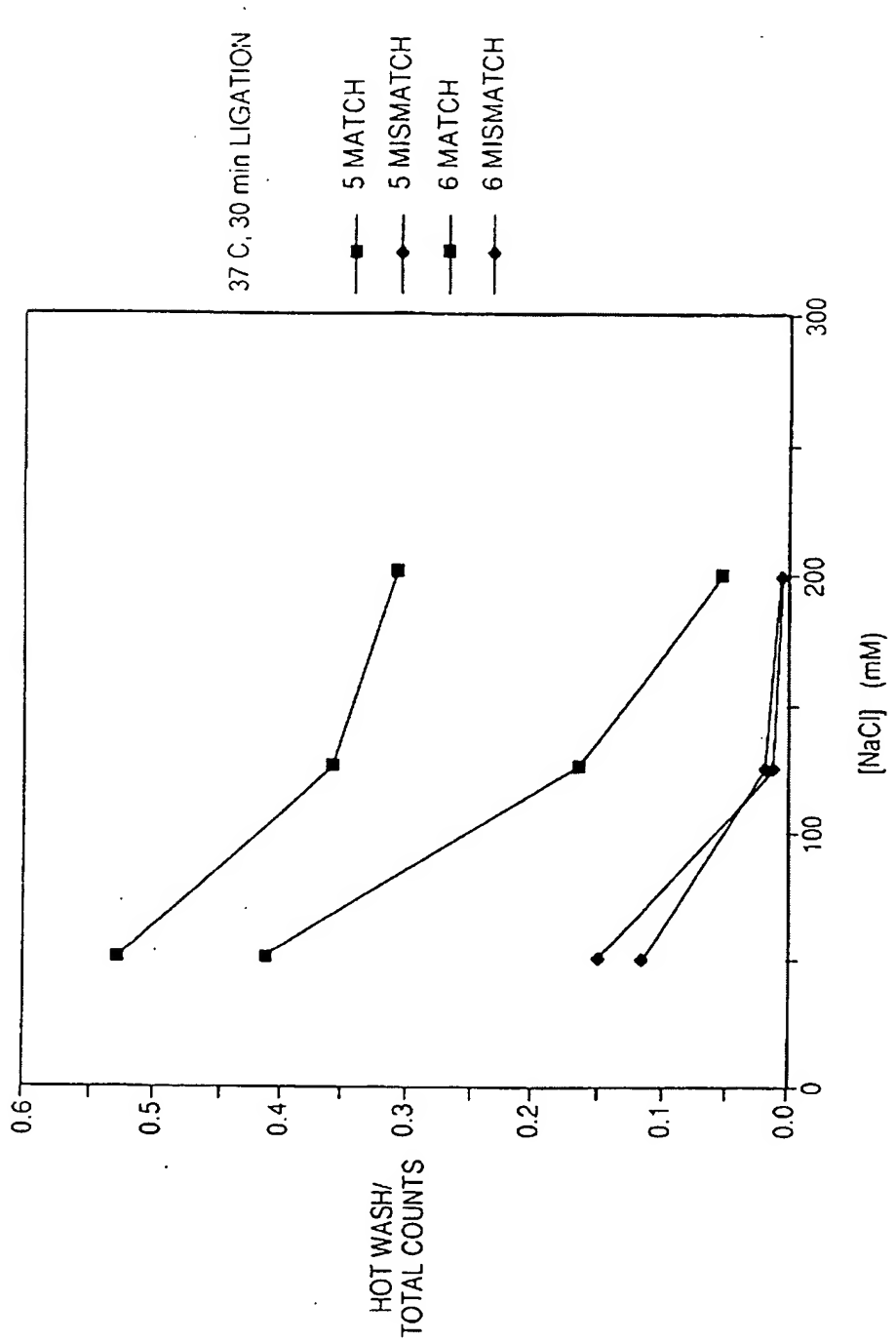
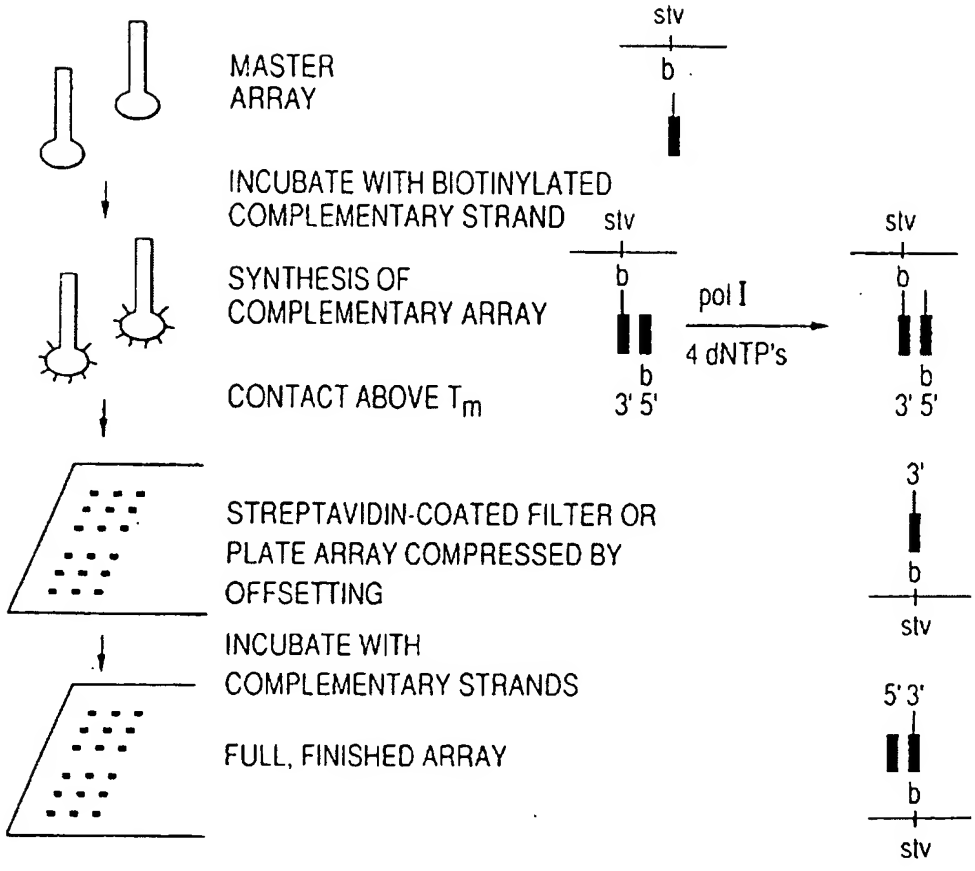

37 C, 30 min LIGATION
5 MATCH
5 MISMATCH
6 MATCH
6 MISMATCH
LIGATION EFFICIENCY
HOT WASH/
TOTAL COUNTS
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0
100
200
300
[NaCl] (mM)

FIG. 13

FIG. 14


MASTER ARRAY
INCUBATE WITH BIOTINYLATED COMPLEMENTARY STRAND
SYNTHESIS OF COMPLEMENTARY ARRAY
CONTACT ABOVE T_m
STREPTAVIDIN-COATED FILTER OR PLATE ARRAY COMPRESSED BY OFFSETTING
INCUBATE WITH COMPLEMENTARY STRANDS
FULL, FINISHED ARRAY
stv
b
pol I
4 dNTP's
3' 5'
5' 3'

METHOD FOR REPLICATING AN ARRAY OF NUCLEIC ACID PROBES

REFERENCE TO RELATED APPLICATIONS

This is a continuation-in-part of U.S. patent application Ser. No. 07/972,012 filed Nov. 6, 1992 now U.S. Pat. No. 5,503,980.

RIGHTS IN THE INVENTION

This invention was made with support from the Department of Energy under grant number PO4591610 and the United States government has certain rights in this invention.

BACKGROUND OF THE INVENTION

1. Field of the Invention

This invention relates to methods for sequencing nucleic acids by positional hybridization and to procedures combining these methods with more conventional sequencing techniques and with other molecular biology techniques including techniques utilized in PCR (polymerase chain reaction) technology. Useful applications include the creation of probes and arrays of probes for detecting, identifying, purifying and sequencing target nucleic acids in biological samples. The invention is also directed to novel methods for the replication of probe arrays, to the replicated arrays, to diagnostic aids comprising nucleic acid probes and arrays useful for screening biological samples for target nucleic acids and nucleic acid variations.

2. Description of the Background

Since the recognition of nucleic acid as the carrier of the genetic code, a great deal of interest has centered around determining the sequence of that code in the many forms which it is found. Two landmark studies made the process of nucleic acid sequencing, at least with DNA, a common and relatively rapid procedure practiced in most laboratories. The first describes a process whereby terminally labeled DNA molecules are chemically cleaved at single base repetitions (A. M. Maxam and W. Gilbert, Proc. Natl. Acad. Sci. USA 74:560–564, 1977). Each base position in the nucleic acid sequence is then determined from the molecular weights of fragments produced by partial cleavages. Individual reactions were devised to cleave preferentially at guanine, at adenine, at cytosine and thymine, and at cytosine alone. When the products of these four reactions are resolved by molecular weight, using, for example, polyacrylamide gel electrophoresis. DNA sequences can be read from the pattern of fragments on the resolved gel.

The second study describes a procedure whereby DNA is sequenced using a variation of the plus-minus method (F. Sanger et al., Proc. Natl. Acad. Sci. USA 74:5463–67, 1977). This procedure takes advantage of the chain terminating ability of dideoxynucleoside triphosphates (ddNTPs) and the ability of DNA polymerase to incorporate ddNTP with nearly equal fidelity as the natural substrate of DNA polymerase, deoxynucleosides triphosphates (dNTPs). A primer, usually an oligonucleotide, and a template DNA are incubated together in the presence of a useful concentration of all four dNTPs plus a limited amount of a single ddNTP. The DNA polymerase occasionally incorporates a dideoxynucleotide which terminates chain extension. Because the dideoxynucleotide has no 3'-hydroxyl, the initiation point for the polymerase enzyme is lost. Polymerization produces a mixture of fragments of varied sizes, all having identical 3' termini. Fractionation of the mixture by, for example, polyacrylamide gel electrophoresis, produces a pattern which indicates the presence and position of each base in the nucleic acid. Reactions with each of the four ddNTPs allows one of ordinary skill to read an entire nucleic acid sequence from a resolved gel.

Despite their advantages, these procedures are cumbersome and impractical when one wishes to obtain megabases of sequence information. Further, these procedures are, for all practical purposes, limited to sequencing DNA. Although variations have developed, it is still not possible using either process to obtain sequence information directly from any other form of nucleic acid.

A new method of sequencing has been developed which overcomes some of the problems associated with current methodologies wherein sequence information is obtained in multiple discrete packages. Instead of having a particular nucleic acid sequenced one base at a time, groups of contiguous bases are determined simultaneously by hybridization. There are many advantages including increased speed, reduced expense and greater accuracy.

Two general approaches of sequencing by hybridization have been suggested. Their practicality has been demonstrated in pilot studies. In one format, a complete set of 4^n nucleotides of length n is immobilized as an ordered array on a solid support and an unknown DNA sequence is hybridized to this array (K. R. Khrapko et al., J. DNA Sequencing and Mapping 1:375–88, 1991). The resulting hybridization pattern provides all n-tuple words in the sequence. This is sufficient to determine short sequences except for simple tandem repeats.

In the second format, an array of immobilized samples is hybridized with one short oligonucleotide at a time (Z. Strezoska et al., Proc. Natl. Acad. Sci. USA 88:10,089–93, 1991). When repeated 4^n times for each oligonucleotide of length n, much of the sequence of all the immobilized samples would be determined. In both approaches, the intrinsic power of the method is that many sequenced regions are determined in parallel. In actual practice the array size is about 10^4 to 10^5.

Another powerful aspect of the method is that information obtained is quite redundant, especially as the size of the nucleic acid probe grows. Mathematical simulations have shown that the method is quite resistant to experimental errors and that far fewer than all probes are necessary to determine reliable sequence data (P. A. Pevzner et al., J. Biomol. Struc. & Dyn. 9:399–410, 1991; W. Bains, Genomics 11:295–301, 1991).

In spite of an overall optimistic outlook, there are still a number of potentially severe drawbacks to actual implementation of sequencing by hybridization. First and foremost among these is that 4^n rapidly becomes quite a large number if chemical synthesis of all of the oligonucleotide probes is actually contemplated. Various schemes of automating this synthesis and compressing the products into a small scale array, a sequencing chip, have been proposed.

A second drawback is the poor level of discrimination between a correctly hybridized, perfectly matched duplexes, and an end mismatch. In part, these drawbacks have been addressed at least to a small degree by the method of continuous stacking hybridization as reported by a Khrapko et al. (FEBS Lett. 256:118–22, 1989). Continuous stacking hybridization is based upon the observation that when a single-stranded oligonucleotide is hybridized adjacent to a double-stranded oligonucleotide, the two duplexes are mutually stabilized as if they are positioned side-to-side due to a stacking contact between them. The stability of the interac-

tion decreases significantly as stacking is disrupted by nucleotide displacement, gap, or terminal mismatch. Internal mismatches are presumably ignorable because their thermodynamic stability is so much less than perfect matches. Although promising, a related problem arises which is the inability to distinguish between weak, but correct duplex formation, and simple background such as non-specific adsorption of probes to the underlying support matrix.

A third drawback is that detection is monochromatic. Separate sequential positive and negative controls must be run to discriminate between a correct hybridization match, a mis-match, and background.

A fourth drawback is that ambiguities develop in reading sequences longer than a few hundred base pairs on account of sequence recurrences. For example, if a sequence the same length of the probe recurs three times in the target, the sequence position cannot be uniquely determined. The locations of these sequence ambiguities are called branch points.

A fifth drawback is the effect of secondary structures in the target nucleic acid. This could lead to blocks of sequences that are unreadable if the secondary structure is more stable than occurs on the complimentary strand.

A final drawback is the possibility that certain probes will have anomalous behavior and for one reason or another, be recalcitrant to hybridization under whatever standard sets of conditions ultimately used. A simple example of this is the difficulty in finding matching conditions for probes rich in G/C content. A more complex example could be sequences with a high propensity to form triple helices. The only way to rigorously explore these possibilities is to carry out extensive hybridization studies with all possible oligonucleotides of length n, under the particular format and conditions chosen. This is clearly impractical if many sets of conditions are involved.

Among the early publication which appeared discussing sequencing by hybridization, E. M. Southern (PCT application no. WO 89/10977, published Nov. 16, 1989; which is hereby specifically incorporated by reference), described methods whereby unknown, or target, nucleic acids are labeled, hybridized to a set of nucleotides of chosen length on a solid support, and the nucleotide sequence of the target determined, at least partially, from knowledge of the sequence of the bound fragments and the pattern of hybridization observed. Although promising, as a practical matter, this method has numerous drawbacks. Probes are entirely single-stranded and binding stability is dependant upon the size of the duplex. However, every additional nucleotide of the probe necessarily increases the size of the array by four fold creating a dichotomy which severly restricts its plausible use. Further, there is an inability to deal with branch point ambiguities or secondary structure of the target, and hybridization conditions will have to be tailored or in some way accounted for for each binding event.

R. Drmanac et al. (U.S. Pat. No. 5,202,231; which is specifically incorporated by reference) is directed to methods for sequencing by hybridization using sets of oligonucleotide probes with randon sequences. These probes, although useful, suffer from some of the same drawbacks as the methodology of Southern (1989), and like Southern, fail to recognize the advantages of stacking interactions.

K. R. Khrapko et al. (FEBS Lett. 256:118–22, 1989; and J. DNA Sequencing and Mapping 1:357–88, 1991) attempt to address some of these problems using a technique referred to as continuous stacking hybridization. With continuous stacking, conceptually, the entire sequence of a target nucleic acid can be determined. Basically, the target is hybridized to an array of probes, again single-stranded, denatured from the array, and the dissociation kinetics of denaturation analyzed to determine the target sequence. Although also promising, discrimination between matches and mis-matches (and simple background) is low, and further, as hybridization conditions are inconstant for each duplex, discrimination becomes increasingly reduced with increasing target complexity.

SUMMARY OF THE INVENTION

The present invention overcomes the problems and disadvantages associated with current strategies and designs and provides new methods for rapidly and accurately determining the nucleotide sequence of a nucleic acid by the herein described methods of positional sequencing by hybridization.

One embodiment of the invention is directed to arrays of R^4 different nucleic acid probes wherein each probe comprises a double-stranded portion of length D, a terminal single-stranded portion of length S, and a random nucleotide sequence within the single-stranded portion of length R. These arrays may be bound to solid supports and are useful for determining the nucleotide sequence of unknown nucleic acids and for the detection, identification and purification of target nucleic acids in biological samples.

Another embodiment of the invention is directed to methods for creating arrays of probes comprising the steps of synthesizing a first set of nucleic acids each comprising a constant sequence of length C at the 3'-terminus, and a random sequence of length R at the 5'-terminus, synthesizing a second set of nucleic acids each comprising a sequence complimentary to the constant sequence of the first nucleic acid, and hybridizing the first set with the second set to form the array.

Another embodiment of the invention is directed to methods for creating arrays of probes comprising the steps of synthesizing a set of nucleic acids each containing a random internal sequence of length R flanked by the cleavage sites of a restriction enzyme, synthesizing a set of primers each compliementary to a non-random sequence of the nucleic acid, hybridizing the two sets together to form hybrids, extending the sequence of the primer by polymerization using the nucleic acid as a template, and cleaving the hybrids with the restriction enzyme to form an array of probes with a double-stranded portion and a single-stranded portion and with the random sequence within the single stranded portion.

Another embodiment of the invention is directed to replicated arrays and methods for replicating arrays of probes, preferably on a solid support, comprising the steps of synthesizing an array of nucleic acids each comprising a constant sequence of length C at a 3'-terminus and a random sequence of length R at a 5'-terminus, fixing the array to a first solid support, synthesizing a set of nucleic acids each comprising a sequence complimentary to the constant region of the array, hybridizing the nucleic acids of the set with the array, enzymatically extending the nucleic acids of the set using the random sequences of the array as templates, denaturing the set of extended nucleic acids, and fixing the denatured nucleic acids of the set to a second solid support to create the replicated array of probes. The replicated array may be single-stranded or double-stranded, it may be fixed to a solid support or free in solution, and it is useful for sequencing, detecting or simply identifying target nucleic acids.

The array is also useful for the purification of nucleic acid from a complex mixture for later identification and/or

sequencing. A purification array comprises sufficient numbers of probes to hybridize and thereby effectively capture the target sequences from a complex sample. The hybridized array is washed to remove non-target nucleic acids and any other materials which may be present and the target sequences eluted by denaturing. From the elution, purified or semi-purified target sequences are obtained and collected. This collection of target sequences can then be subjected to normal sequencing methods or sequenced by the methods described herein.

Another embodiment of the invention is directed to nucleic acid probes and methods for creating nucleic acid probes comprising the steps of synthesizing a plurality of single-stranded first nucleic acids and a plurality of longer single-stranded second nucleic acids wherein each each second nucleic acid comprises a random terminal sequence and a sequence complimentary to a sequence of the first nucleic acids, hybridizing the first nucleic acids to the second to form partial duplexes having a double-stranded portion and a single-stranded portion with the random sequence within the single-stranded portion, hybridizing a target nucleic acid to the partial duplexes, optionally ligating the hybridized target to the first nucleic acid of the partial duplexes, isolating the second nucleic acid from the ligated duplexes, synthesizing a plurality of third nucleic acids each complimentary to the constant sequence of the second nucleic acid, and hybridizing the third nucleic acids with the isolated second nucleic acids to create the nucleic acid probe. Alternatively, after formation of the partial duplexes, the target is ligated as before and hybridized with a set of oligonucleotides comprising random sequences. These oligonucleotides are ligated to the second nucleic acid, the second nucleic acid is isolated, another plurality of first nucleic acids are synthesized, and the first nucleic acids are hybridized to the oligonucleotide ligated second nucleic acids to form the probe. Ligation allows for hybridization to be performed under a single set of hybridization conditions. Probes may be fixed to a solid support and may also contain enzyme recognition sites within their sequences.

Another embodiment of the invention is directed to diagnostic aids and methods utilizing probe arrays for the detection and identification of target nucleic acids in biological samples and to methods for using the diagnostic aids to screen biological samples. Diagnostic aids as described are also useful for the purification of identified targets and, if desired, for their sequencing. These aids comprise probes, solid supports, labels, necessary reagents and the biological samples.

Other advantages of the invention are set forth in part in the description which follows, and in part, will be obvious from this description, or may be learned from the practice of this invention. The accompanying drawings which are incorporated in and constitute a part of this specification, illustrate and, together with this description, serve to explain the principle of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 Energetics of stacking hybridization. Structures consist of a long target and a probe of length n. The top three sample are ordinary hybridization and the bottom three are stacking hybridization.

FIG. 2 (A) The first step of the basic scheme for positional sequencing by hybridization depicting the hybridization of target nucleic acid with probe forming a 5' overhang of the target.

(B) The first step of the alternate scheme for positional sequencing by hybridization depicting the hybridization of target nucleic acid with probe forming a 3' overhang of the probe.

FIG. 3 Graphic representation of the ligation step of positional sequencing by hybridization wherein hybridization of the target nucleic acid produces (A) a 5' overhang or (B) a 3' overhang.

FIG. 4 Preparation of a random probe array.

FIG. 5 Single nucleotide extension of a probe hybridized with a target nucleic acid using DNA polymerase and a single dideoxynucleotide.

FIG. 6 Preparation of a nested set of targets using labeled target nucleic acids partially digested with exonuclease III.

FIG. 7 Determination of positional information using the ratio of internal label to terminal label.

FIG. 8 (A) Extension of one strand of the probe using a hybridized target as template with a single deoxynucleotide.

(B) Hybridization of target with a fixed probe followed by ligation of probe to target.

FIG. 9 Four color analysis of sequence extensions of the 3' end of a probe using three labeled nucleoside triphosphates and one unlabeled chain terminator.

FIG. 10 Extension of a nucleic acid probe by ligation of a pentanucleotide 3' blocked to prevent polymerization.

FIG. 11 Preparation of a customized probe containing a 10 base pair sequence that was present in the original target nucleic acid.

FIG. 12 Graphic representation of the general procedure of positional sequencing by hybridization.

FIG. 13 Graphical representation of the ligation efficiency of positional sequencing. Depicted is the relationship between the amount of label remaining over the total amounts of label in the reaction, verses NaCl concentration.

FIG. 14 A diagrammatic representation of the construction of a complimentary array of master beads.

DESCRIPTION OF THE INVENTION

The present invention overcomes the problems and disadvantages associated with current strategies and designs and provides new methods and probes, new diagnostic aids and methods for using the diagnostic aids, and new arrays and methods for creating arrays of probes to detect, identify, purify and sequence target nucleic acids. Nucleic acids of the invention include sequences of deoxyribonucleic acid (DNA) or ribonucleic acid (RNA) which may be isolated from natural sources, recombinantly produced, or artificially synthesized. Preferred embodiments of the present invention is probe synthesized using traditional chemical synthesis, using the more rapid polymerase chain reaction (PCR) technology, or using a combination of these two methods.

Nucleic acids of the invention further encompass polyamide nucleic acid (PNA) or any sequence of what are commonly referred to as bases joined by a chemical backbone that have the ability to base pair or hybridize with a complimentary chemical structure. The bases of DNA, RNA, and PNA are purines and pyrimidines linearly linked to a chemical backbone. Common chemical backbone structures are deoxyribose phosphate and ribose phosphate. Recent studies demonstrated that a number of additional structures may also be effective, such as the polyamide backbone of PNA (P. E. Nielsen et al., Sci. 254:1497–1500, 1991).

The purines found in both DNA and RNA are adenine and guanine, but others known to exist are xanthine, hypoxanthine, 2- and 1-diaminopurine, and other more modified bases. The pyrimidines are cytosine, which is common to both DNA and RNA, uracil found predomi-

nantly in RNA, and thymidine which occurs exclusively in DNA. Some of the more atypical pyrimidines include methylcytosine, hydroxymethyl-cytosine, methyluracil, hydroxymethyluracil, dihydroxypentyluracil, and other base modifications. These bases interact in a complimentary fashion to form base-pairs, such as, for example, guanine with cytosine and adenine with thymidine. However, this invention also encompasses situations in which there is nontraditional base pairing such as Hoogsteen base pairing which has been identified in certain tRNA molecules and postulated to exist in a triple helix.

One embodiment of the invention is directed to a method for determining a nucleotide sequence by positional hybridization comprising the steps of (a) creating a set of nucleic acid probes wherein each probe has a double-stranded portion, a single-stranded portion, and a random sequence within the single-stranded portion which is determinable, (b) hybridizing a nucleic acid target which is at least partly single-stranded to the set of nucleic acid probes, and (c) determining the nucleotide sequence of the target which hybridized to the single-stranded portion of any probe. The set of nucleic acid probes and the target nucleic acid may comprise DNA, RNA, PNA, or any combination thereof, and may be derived from natural sources, recombinant sources, or be synthetically produced. Each probe of the set of nucleic acid probes has a double-stranded portion which is preferably about 10 to 30 nucleotides in length, a single-stranded portion which is preferably about 4 to 20 nucleotides in length, and a random sequence within the single-stranded portion which is preferably about 4 to 20 nucleotides in length and more preferably about 5 nucleotides in length. A principle advantage of this probe is in its structure. Hybridization of the target nucleic acid is encouraged due to the favorable thermodynamic conditions established by the presence of the adjacent double-strandedness of the probe. An entire set of probes contains at least one example of every possible random nucleotide sequence.

By way of example only, if the random portion consisted of a four nucleotide sequence (R=4) of adenine, guanine, thymine, and cystosine, the total number of possible combinations (4^R) would be 4^4 or 256 different nucleic acid probes. If the number of nucleotides in the random sequence was five, the number of different probes within the set would be 4^5 or 1,024. This becomes a very large number indeed when considering sequences of 20 nucleotides or more.

However, to determine the complete sequence of a nucleic acid target, the set of probes need not contain every possible combination of nucleotides of the random sequence to be encompassed by the method of this invention. This variation of the invention is based on the theory of degenerated probes proposed by S. C. Macevicz (International Patent Application, US89-04741, published 1989, and herein specifically incorporated by reference). The probes are divided into four subsets. In each, one of the four bases is used at a defined number of positions and all other bases except that one on the remaining positions. Probes from the first subset contain two elements, A and non-A (A=adenosine). For a nucleic acid sequence of length k, there are $4(2^k-1)$, instead of 4^k probes. Where k=8, a set of probes would consist of only 1020 different members instead of the entire set of 65,536. The savings in time and expense would be considerable. In addition, it is also a method of the present invention to utilize probes wherein the random nucleotide sequence contains gapped segments, or positions along the random sequence which will base pair with any nucleotide or at least not interfere with adjacent base pairing.

Hybridization between complimentary bases of DNA, RNA, PNA, or combinations of DNA, RNA and PNA, occurs under a wide variety of conditions such as variations in temperature, salt concentration, electrostatic strength, and buffer composition. Examples of these conditions and methods for applying them are described in *Nucleic Acid Hybridization: A Practical Approach* (B. D. Hames and S. J. Higgins, editors, IRL Press, 1985), which is herein specifically incorporated by reference. It is preferred that hybridization takes place between about 0° C. and about 70° C., for periods of from about 5 minutes to hours, depending on the nature of the sequence to be hybridized and its length. For example, typical hybridization conditions for a mixture of two 20-mers is to bring the mixture to 68° C. and let cool to room temperature (22° C.) for five minutes or at very low temperatures such as 2° C. in 2 microliters. It is also preferred that hybridization between nucleic acids be facilitated using buffers such as saline, Tris-EDTA (TE), Tris-HCl and other aqueous solutions, certain reagents and chemicals. Preferred examples of these reagents include single-stranded binding proteins such as Rec A protein, T4 gene 32 protein, *E. coli* single-stranded binding protein, and major or minor nucleic acid groove binding proteins. Preferred examples of other reagents and chemicals include divalent ions, polyvalent ions, and intercalating substances such as ethidium bromide, actinomycin D, psoralen, and angelicin.

The nucleotide sequence of the random portion of each probe is determinable by methods which are well-known in the art. Two methods for determining the sequence of the nucleic acid probe are by chemical cleavage, as disclosed by Maxam and Gilbert (1977), and by chain extension using ddNTPs, as disclosed by Sanger et al. (1977), both of which are herein specifically incorporated by reference. Alternatively, another method for determining the nucleotide sequence of a probe is to individually synthesize each member of a probe set. The entire set would comprise every possible sequence within the random portion or some smaller portion of the set. The method of the present invention could then be conducted with each member of the set. Another procedure would be to synthesize one or more sets of nucleic acid probes simultaneously on a solid support. Preferred examples of a solid support include a plastic, a ceramic, a metal, a resin, a gel, and a membrane. A more preferred embodiment comprises a two-dimensional or three-dimensional matrix, such as a gel, with multiple probe binding sites, such as a hybridization chip as described by Pevzner et al. (J. Biomol. Struc. & Dyn. 9:399–410, 1991), and by Maskos and Southern (Nuc. Acids Res. 20:1679–84, 1992), both of which are herein specifically incorporated by reference. Nucleic acids are bound to the solid support by covalent binding such as by conjugation with a coupling agent, or by non-covalent binding such as an electrostatic interaction or antibody-antigen coupling. Typical coupling agents include biotin/streptavidin, *Staphylococcus aureus* protein A/IgG antibody F_c fragment, and streptavidin/protein A chimeras (T. Sano and C. R. Cantor, Bio/Technology 9:1378–81, 1991).

Hybridization chips can be used to construct very large probe arrays which are subsequently hybridized with a target nucleic acid. Analysis of the hybridization pattern of the chip provides an immediate fingerprint identification of the target nucleotide sequence. Patterns can be manually or computer analyzed, but it is clear that positional sequencing by hybridization lends itself to computer analysis and automation. Algorithms and software have been developed for sequence reconstruction which are applicable to the methods described herein (R. Drmanac et al., J. Biomol. Struc. & Dyn. 5:1085–1102, 1991; P. A. Pevzner, J. Biomol. Struc. & Dyn. 7:63–73, 1989, both of which are herein specifically incorporated by reference).

Preferably, target nucleic acids are labeled with a detectable label. Label may be incorporated at a 5' terminal site, a 3' terminal site, or at an internal site within the length of the nucleic acid. Preferred detectable labels include a radioisotope, a stable isotope, an enzyme, a fluorescent chemical, a luminescent chemical, a chromatic chemical, a metal, an electric charge, or a spatial structure. There are many procedures whereby one of ordinary skill can incorporate detectable label into a nucleic acid. For example, enzymes used in molecular biology will incorporate radioisotope labeled substrate into nucleic acid. These include polymerases, kinases, and transferases. The labeling isotope is preferably, ^{32}P, ^{35}S, ^{14}C, or ^{125}I.

Label may be directly or indirectly detected using scintillation fluid or a PhosphorImager, chromatic or fluorescent labeling, or mass spectrometry. Other, more advanced methods of detection include evanescent wave detection of surface plasmon resonance of thin metal film labels such as gold, by, for example, the BIAcore sensor sold by Pharmacia, or other suitable biosensors. Alternatively, the probe may be labeled and the target nucleic acid detected, identified and possibly sequenced from interaction with the labeled probe. For example, a labeled probe or array of probes may be fixed to a solid support. From an analysis of the binding observed after hybridization with a biological sample containing nucleic acid, the target nucleic acid is identified.

Another embodiment of the invention is directed to methods for determining a sequence of a nucleic acid comprising the steps of labeling the nucleic acid with a first detectable label at a terminal site, labeling the nucleic acid with a second detectable label at an internal site, identifying the nucleotide sequences of portions of the nucleic acid, determining the relationship of the nucleotide sequence portions to the nucleic acid by comparing the first detectable label and the second detectable label, and determining the nucleotide sequence of the nucleic acid. Fragments of target nucleic acids labeled both terminally and internally can be distinguished based on the relative amounts of each label within respective fragments. Fragments of a target nucleic acid terminally labeled with a first detectable label will have the same amount of label as fragments which include the labeled terminus. However, theses fragments will have variable amounts of the internal label directly proportional to their size and distance for the terminus. By comparing the relative amount of the first label to the relative amount of the second label in each fragment, one of ordinary skill is able to determine the position of the fragment or the position of the nucleotide sequence of that fragment within the whole nucleic acid.

Another embodiment of the invention is directed to methods for determining a nucleotide sequence by hybridization comprising the steps of creating a set of nucleic acid probes wherein each probe has a double-stranded portion, a single-stranded portion, and a random sequence within the single-stranded portion which is determinable, hybridizing a nucleic acid target which is at least party single-stranded to the set, ligating the hybridized target to the probe, and determining the nucleic sequence of the target which is hybridized to the single-stranded portion of any probe. This embodiment adds a step wherein the hybridized target is ligated to the probe. Ligation of the target nucleic acid to the complimentary probe increases fidelity of hybridization and allows for incorrectly hybridized target to be easily washed from correctly hybridized target (FIG. 11). More importantly, the addition of a ligation step allows for hybridization to be performed under a single set of hybridization conditions. For example, hybridization temperature is preferably between about 22°–37° C., the salt concentration useful is preferably between about 0.05–0.5M, and the period of hybridization is between about 1–14 hours. This is not possible using the methodologies of the current procedures which do not employ a ligation step and represents a very substantial improvement. Ligation can be accomplished using a eukaryotic derived or a prokaryotic derived ligase. Preferred is T4 DNA or RNA ligase. Methods for use of these and other nucleic acid modifying enzymes are described in *Current Protocols in Molecular Biology* (F. M. Ausubel et al., editors, John Wiley & Sons, 1989), which is herein specifically incorporated by reference.

There are a number of distinct advantages to the incorporation of a ligation step. First and foremost is that one can use identical hybridization conditions for hybridization. Variation of hybridization conditions due to base composition are no longer relevant as nucleic acids with high A/T or G/C content ligate with equal efficiency. Consequently, discrimination is very high between matches and mismatches, much higher than has been achieved using other methodologies such as Southern (1989) wherein the effects of G/C content were only somewhat neutralized in high concentrations of quarternary or tertiary amines (e.g., 3M tetramethyl ammonium chloride in Drmanac et al., 1993).

Another embodiment of the invention is directed to methods for determining a nucleotide sequence by hybridization which comprises the steps of creating a set of nucleic acid probes wherein each probe has a double-stranded portion, a single-stranded portion, and a random sequence within the single-stranded portion which is determinable, hybridizing a target nucleic acid which is at least partly single-stranded to the set of nucleic acid probes, enzymatically extending a strand of the probe using the hybridized target as a template, and determining the nucleotide sequence of the single-stranded portion of the target nucleic acid. This embodiment of the invention is similar to the previous embodiment, as broadly described herein, and includes all of the aspects and advantages described therein. An alternative embodiment also includes a step wherein hybridized target is ligated to the probe. Ligation increases the fidelity of the hybridization and allows for a more stringent wash step wherein incorrectly hybridized, unligated target can be removed and further, allows for a single set of hybridization conditions to be employed. Most nonligation techniques including Southern (1989), Drmanac et al. (1993), and Khrapko et al. (1989 and 1991), are only accurate, and only marginally so, when hybriizations are performed under optimal conditions which vary with the G/C content of each interaction. Preferable condiions comprise a hybridization temperature of between about 22°–37° C., a salt concentration of betwen about 0.05–0.5M, and a hybridization period of between about 1–14 hours.

Hybridization produces either a 5' overhang or a 3' overhang of target nucleic acid. Where there is a 5' overhang, a 3-hydroxyl is available on one strand of the probe from which nucleotide addition can be initiated. Preferred enzymes for this process include eukaryotic or prokaryotic polymerases such as T3 or T7 polymerase, Klenow fragment, or Taq polymerase. Each of these enzymes are readily available to those of ordinary skill in the art as are procedures for their use (*Current Protocols in Molecular Biology*).

Hybridized probes may also be enzymatically extended a predetermined length. For example, reaction condition can be established wherein a single dNTP or ddNTP is utilized as substrate. Only hybridized probes wherein the first nucle-

otide to be incorporated is complimentary to the target sequence will be extended, thus, providing additional hybridization fidelity and additional information regarding the nucleotide sequence of the target. Sanger (1977) or Maxam and Gilbert (1977) sequencing can be performed which would provide further target sequence data. Alternatively, hybridization of target to probe can produces 3' extensions of target nucleic acids. Hybridized probes can be extended using nucleoside biphosphate substrates or short sequences which are ligated to the 5' terminus.

Another embodiment of the invention is directed to a method for determining a nucleotide sequence of a target by hybridization comprising the steps of creating a set of nucleic acid probes wherein each probe has a double-stranded portion, a single-stranded portion, and a random nucleotide sequence within the single-stranded portion which is determinable, cleaving a plurality of nucleic acid targets to form fragments of various lengths which are at least partly single-stranded, hybridizing the single-stranded region of the fragments with the single-stranded region of the probes, identifying the nucleotide sequences of the hybridized portions of the fragments, and comparing the identified nucleotide sequences to determine the nucleotide sequence of the target. An alternative embodiment includes a further step wherein the hybridized fragments are ligated to the probes prior to identifying the nucleotide sequences of the hybridized portions of the fragments. As described herein, the addition of a ligation step allows for hybridizations to be performed under a single set of hybridization conditions.

In these embodiments, target nucleic acid is partially cleaved forming a plurality of nucleic acid fragments of various lengths, a nested set, which is then hybridized to the probe. It is preferred that cleavage occurs by enzymatic, chemical or physical means. Preferred enzymes for partial cleavage are exonuclease III, S1 nuclease, DNase I, Bal 31, mung bean nuclease, P1 nuclease, lambda exonuclease, restriction endonuclease, and RNase I. Preferred means for chemical cleavage are ultraviolet light induced cleavage, ethidium bromide induced cleavage, and cleavage induced with acid or base. Preferred means for mechanical cleavage are shearing through direct agitation such as vortexing or multiple cycles of freeze-thawing. Procedures for enzymatic, chemical or physical cleavage are disclosed in, for example, *Molecular Cloning: A Laboratory Manual* (T. Maniatis et al., editors, Cold Spring Harbor 1989), which is herein specifically incorporated by reference.

Fragmented target nucleic acids will have a distribution of terminal sequences which is sufficiently broad so that the nucleotide sequence of the hybridized fragments will include the entire sequence of the target nucleic acid. A preferred method is wherein the set of nucleic acid probes is fixed to a solid support. A preferred solid support is a plastic, a ceramic, a metal or magnetic substance, a resin, a film or other polymer, a gel, or a membrane, and it is more preferred that the solid support be a two-dimensional or three-dimensional matrix with multiple probe binding sites such as a hybridization chip as described by K. R. Khrapko et al. (J. DNA Sequencing and Mapping 1:357–88, 1991). It is also preferred wherein the target nucleic acid has a detectable label such as a radioisotope, a stable isotope, an enzyme, a fluorescent chemical, a luminescent chemical, a chromatic chemical, a metal, an electric charge, or a spatial structure.

As an extension of this procedure, it is also possible to use the methods herein described to determine the nucleotide sequence of one or more probes which hybridize with an unknown target sequence. For example, fragmented targets could be terminally or internally labeled, hybridized with a set of nucleic acid probes, and the hybridized sequences of the probes determined. This aspect may be useful when it is cumbersome to determine the sequence of the entire target and only a smaller region of that sequence is of interest.

Another embodiment of the invention is directed a method wherein the target nucleic acid has a first detectable label at a terminal site and a second detectable label at an internal site. The labels may be the same type of label or of different types as long as each can be discriminated, preferably by the same detection method. It is preferred that the first and second detectable labels are chromatic or fluorescent chemicals or molecules which are detectable by mass spectrometry. Using a double-labeling method coupled with analysis by mass spectrometry provides a very rapid and accurate sequencing methodology that can be incorporated in sequencing by hybridization and lends itself very well to automation and computer control.

Another embodiment of the invention is directed to methods for creating a nucleic acid probe comprising the steps of synthesizing a plurality of single-stranded first nucleic acids and an array of longer single-stranded second nucleic acids complimentary to the first nucleic acid with a random terminal nucleotide sequence, hybridizing the first nucleic acids to the second nucleic acids to form hybrids having a double-stranded portion and a single-stranded portion with the random nucleotide sequence within the single-stranded portion, hybridizing a single-stranded nucleic acid target to the hybrids, ligating the hybridized target to the first nucleic acid of the hybrid, isolating the second nucleic acid, and hybridizing the first nucleic acid of step with the isolated second nucleic acid to form a nucleic acid probe. Probes created in this manner are referred to herein as customized probes.

Preferred customized probe comprises a first nucleic acid which is about 15–25 nucleotides in length and the second nucleic acid is about 20–30 nucleotides in length. It is also preferred that the double-stranded portion contain an enzyme recognition site which allows for increased flexibility of use and facilitates cloning, should it at some point become desirable to clone one or more of the probes. It is also preferred if the customized probe is fixed to a solid support, such as, a plastic, a ceramic, a metal, a resin, a film or other polymer, a gel, or a membrane, or possibly a two- or three-dimensional array such as a chip or microchip.

Customized probes, created by the method of this invention, have a wide range of uses. These probes are, first of all, structurally useful for identifying and binding to only those sequences which are homologous to the overhangs. Secondly, the overhangs of these probes possess the nucleotide sequence of interest. No further manipulation is required to carry the sequence of interest to another structure. Therefore, the customized probes greatly lend themselves to use in, for example, diagnostic aids for the genetic screening of a biological sample.

Another embodiment of the invention is directed to arrays of nucleic acid probes wherein each probe comprises a double-stranded portion of length D, a terminal single-stranded portion of length S, and a random nucleotide sequence within the single-stranded portion of length R. Preferably, D is between about 3–20 nucleotides and S is between about 3–20 nucleotides and the entire array is fixed to a solid support which may be composed of plastics, ceramics, metals, resins, polymers and other films, gels, membranes and two-dimensional and three-dimensional matrices such as hybridization chips or microchips. Probe

arrays are useful in sequencing and diagnostic applications when the sequence and/or position on a solid support of every probe of the array is known or is unknown. In either case, information about the target nucleic acid may be obtained and the target nucleic acid detected, identified and sequenced as described in the methods described herein. Arrays comprise 4^R different probes representing every member of the random sequence of length R, but arrays of less than 4^R are also encompassed by the invention.

Another embodiment of the invention is directed to method for creating probe arrays comprising the steps of synthesizing a first set of nucleic acids each comprising a constant sequence of length C at a 3'-terminus and a random sequence of length R at a 5'-terminus, synthesizing a second set of nucleic acids each comprising a sequence complimentary to the constant sequence of each of the first nucleic acid, and hybridizing the first set with the second set to create the array. Preferably, the nucleic acids of the first set are each between about 15–30 nucleotides in length and the nucleic acids of the second set are each between about 10–25 nucleotides in length. Also preferable is that C is between about 7–20 nucleotides and R is between about 3–10 nucleotides.

Arrays may comprise about 4^R different probes, but in certain applications, an entire array of every possible sequence is not necessary and incomplete arrays are acceptable for use. For example, incomplete arrays may be utilized for screening procedures of very rare target nucleic acids where nonspecific hybridization is not expected to be problematic. Further, every member of an array may not be needed when detecting or sequencing smaller nucleic acids where the chance of requiring certain combinations of nucleotides is so low as to be practically nonexistent. Array which are fixed to solid supports are expected to be most useful, although array in solution also have many applications. Solid supports which are useful include plastics such as microtiter plates, beads and microbeads, ceramics, metals where resilience is desired or magnetic beads for ease of isolation, resins, gels, polymers and other films, membranes or chips such as the two- and three-dimensional sequencing chips utilized in sequencing technology.

Alternatively, probe arrays may also be made which are single-stranded. These arrays are created, preferably on a solid support, basically as described, by synthesizing an array of nucleic acids each comprising a constant sequence of length C at a 3'-terminus and a random sequence of length R at a 5'-terminus, and fixing the array to a first solid support. Arrays created in this manner can be quickly and easily transformed into double-stranded arrays by the synthesis and hybridization of a set of nucleic acids with a sequence complimentary to the constant sequence of the replicated array to create a double-stranded replicated array. However, in their present form, single-stranded arrays are very valuable as templates for replication of the array.

Due to the very large numbers of probes which comprise most useful arrays, there is a great deal of time spent in simply creating the array. It requires many hours of nucleic acid synthesis to create each member of the array and many hours of manipulations to place the array in an organized fashion onto any solid support such as those described previously. Once the master array is created, replicated arrays or slaves, can be quickly and easily created by the methods of the invention which take advantage of the speed and accuracy of nucleic acid polymerases. Basically, methods for replicating an array of single-stranded probes on a solid support comprise the steps of synthesizing an array of nucleic acids each comprising a constant sequence of length C at a 3'-terminus and a random sequence of length R at a 5'-terminus, fixing the array to a first solid support, synthesizing a set of nucleic acids each comprising a sequence complimentary to the constant sequence, hybridizing the nucleic acids of the set with the array, enzymatically extending the nucleic acids of the set using the random sequences of the array as templates, denaturing the set of extended nucleic acids, and fixing the denatured nucleic acids of the set to a second solid support to create the replicated array of single-stranded probes.

Denaturation of the array can be performed by subjecting the array to heat, for example 90°–100° C. for 2–15 minutes, or highly alkaline conditions, such as by the addition of sodium hydroxide. Denaturation can also be accomplished by adding organic solvents, nucleic acid binding proteins or enzymes which promote denaturation to the array. Preferably, the solid supports are coated with a substance such as streptavidin and the nucleic acid reagents conjugated with biotin. Denaturation of the partial duplex leads to binding of the nucleic acids to the solid support.

Another embodiment of the invention is directed to methods for creating arrays of probes comprising the steps of synthesizing an array of single-stranded nucleic acids each containing a constant sequence at the 3'-terminus, another constant sequence at the 5'-terminus, and a random internal sequence of length R flanked by the cleavage site(s) of a restriction enzyme (on one or both sides), synthesizing an array of primers each complementary to a portion of the constant sequence of the 3'-terminus, hybridizing the two arrays together to form hybrids, extending the sequence of each primer by polymerization using a sequence of the nucleic acid as a template, and cleaving the extended hybrids with the restriction enzyme to form an array of probes with a double-stranded portion at one terminus, a single-stranded portion containing the random sequence at the opposite terminus. Preferably, the nucleic acids are each between about 10–50 nucleotides in length and R is between about 3–5 nucleotides in length. Any of the restriction enzymes which produce a 3'- or 5'-overhang after cleavage are suitable for use to make the array. Some of the restriction enzymes which are useful in this regard, and their recognition sequences are depicted in Table 1.

TABLE 1

Restriction Enzyme	Recognition Sequence 5'-Overhang	3'-Overhang
AlwN I	5'-CAG NNN↓CTG	
	3'-GTC↑NNN GAC	
Bbv I	5'-GCAGC(N)$_8$↓ (SEQ ID NO 37)	
	3'-CGTCG(N)$_{12}$↑ (SEQ ID NO 38)	
Bgl I	5'-GCCN NNN↓NGGC (SEQ ID NO 39)	

TABLE 1-continued

Restriction Enzyme	Recognition Sequence 5'-Overhang	3'-Overhang
	3'CGGN↑NNN NCCG (SEQ ID NO 39)	
BstX I		5'-CCAN NNNN↓NTGG (SEQ ID NO 40) 3'-GGTN↑NNNN NACC (SEQ ID NO 40)
Dra III	5'-CAC NNN↓GTG 3'-GTG↑NNN CAC	
Fok I	5'-GGATG(N)$_9$↓ (SEQ ID NO 41) 3'-CCTAC(N)$_{13}$↑ (SEQ ID NO 42)	
Hga I	5'-GACGC(N)$_5$↓ (SEQ ID NO 43) 3'-CTGCG(N)$_{10}$↑ (SEQ ID NO 44)	
PflM I		5'-CCAN NNN↓NTGG (SEQ ID NO 45) 3'-GGTN↑NNN NACC (SEQ ID NO 45)
SfaN I	5'-GCATC(N)$_5$↓ (SEQ ID NO 46) 3'-CGTAG(N)$_9$↑ (SEQ ID NO 47)	
Dfi I		5'-GGCCN NNN↓NGGCC (SEQ ID NO 48) 3'-CCGGN↑NNN NCCGG (SEQ ID NO 48)

Also prefered is that the array be fixed to a solid support such as a plastic, ceramic, metal, resin, polymer, gel, film, membrane or chip. Fixation can be accomplished by conjugating the reagents for synthesis with a specific binding protein or other similar substance and coating the surface of the support with the binding counterpart (e.g. biotin/streptavidin, F$_c$/protein A, nucleic acid/nucleic acid binding protein).

Alternatively, another similar method for creating an array of probes comprising the steps of synthesizing an array of single-stranded nucleic acids each containing a constant sequence at the 3'-terminus, another constant sequence at the 5'-terminus, and a random internal sequence of length R flanked by the cleavage site(s) of a restriction enzyme (on one or both sides), synthesizing an array of primers with a sequence complimentary to the constant sequence at the 3'-terminus, hybridizing the two arrays together to form hybrids, enzymatically extending the primers using the nucleic acids as templates to form full-length hybrids, cloning the full-length hybrids into vectors such as plasmids or phage, cloning the plasmids into competent bacteria or phage, reisolating the cloned plasmid DNA, amplifying the cloned sequences by multiple polymerase chain reactions, and cleaving the amplified sequences with the restriction enzyme to form the array of probes with a double-stranded portion at one terminus and a single-stranded portion containing the random sequence at the opposite terminus. Using this method the array of probes may have 5'- or 3'-overhangs depending on the cleavage specificity of the restriction enzyme (e.g. Table 1). The array of probes may be fixed to a solid support such as a plastic, ceramic, metal, resin, polymer, film, gel, membranes and chip. Preferably, during PCR amplification, the reagent primers are conjugated with biotin which facilitates eventual binding to a streptavidin coated surface.

Another embodiment of the invention is directed to methods for using customized probes, arrays, and replicated arrays, as described herein, in diagnostic aids to screen biological samples for specific nucleic acid sequences. Diagnostic aids and methods for using diagnostic aids would be very useful when sequence information at a particular locus of, for example, DNA is desired. Single nucleotide mutations or more complex nucleic acid fingerprints can be identified and analyzed quickly, efficiently, and easily. Such an approach would be immediately useful for the detection of individual and family genetic variation, of inherited mutations such as those which cause a disease, DNA dependent normal phenotypic variation, DNA dependent somatic variation, and the presence of heterologous nucleic acid sequences.

Especially useful are diagnostic aids comprising probe arrays. These arrays can make the detection identification, and sequencing of nucleic acids from biological samples exceptionally rapid and allows one to obtain multiple pieces of information from a single sample after performing a single test. Methods for detecting and/or identifying a target nucleic acid in a biological sample comprise the steps of creating an array of probes fixed to a solid support as described herein, labeling the nucleic acid of the biological sample with a detectable label, hybridizing the labeled nucleic acid to the array and detecting the sequence of the nucleic acid from a binding pattern of the label on the array. These methods for creating probe arrays and for rapidly and efficiently replicating those arrays, such as for diagnostic aids, makes the manufacture and commercial application of large numbers of arrays a possibility.

As described, these diagnostic aids are useful to humans, other animals, and even plants for the detection of infections due to viruses, bacteria, fungi or yeast, and for the detection of certain parasites. These detection methods and aids are also useful in the feed and food industries and in the environmental field for the detection, identification and sequencing of nucleic acids associated with samples obtained from environmental sources and from manufacturing products and by-products.

Diagnostic aids comprise specific nucleic acid probes fixed to a solid support to which is added the biological sample. Hybridization of target nucleic acids is determined by adding a detectable label, such as a labeled antibody, which will specifically recognize only hybridized targets or, alternatively, unhybridized target is washed off and labeled target specific antibodies are added. In either case, appearance of label on the solid support indicates the presence of nucleic acid target hybridized to the probe and consequently, within the biological sample.

Customized probes may also prove useful in prophylaxis or therapy by directing a drug, antigen, or other substance to a nucleic acid target with which it will hybridize. The substance to be targeted can be bound to the probe so as not to interfere with possible hybridization. For example, if the probe was targeted to a viral nucleic acid target, an effective antiviral could be bound to the probe which will then be able to specifically carry the antiviral to infected cells. This

would be especially useful when the treatment is harmful to normal cells and precise targeting is required for efficacy.

Another embodiment of the invention is directed to methods for creating a nucleic acid probe comprising the steps of synthesizing a plurality of single-stranded first nucleic acids and an array of longer single-stranded second nucleic acids complimentary to the first nucleic acid with a random terminal nucleotide sequence, hybridizing the first nucleic acids to the second nucleic acids to form hybrids having a double-stranded portion and a single-stranded portion with the random nucleotide sequence within the single-stranded portion, hybridizing a single-stranded nucleic acid target to the hybrids, ligating the hybridized target to the first nucleic acid of the hybrid, hybridizing the ligated hybrid with an array of oligonucleotides with random nucleotide sequences, ligating the hybridized oligonucleotide to the second nucleic acid of the ligated hybrid, isolating the second nucleic acid, and hybridizing another first nucleic acid with the isolated second nucleic acid to form a nucleic acid probe. Preferred is that the first nucleic acid is about 15–25 nucleotides in length, that the second nucleic acid is about 20–30 nucleotides in length, that the constant portion contain an enzyme recognition site, and that the oligonucleotides are each about 4–20 nucleotides in length. Probes may be fixed to a solid support such as a plastic, ceramic, a metal, a resin, a gel, or a membrane. It is preferred that the solid support be a two-dimensional or three-dimensional matrix with multiple probe binding sites such as a hybridization chip. Nucleic acid probes created by the method of the present invention are useful in a diagnostic aid to screen a biological sample for genetic variations of nucleic acid sequences therein.

Another embodiment of the invention is directed to a method for creating a nucleic acid probe comprising the steps of (a) synthesizing a plurality of single-stranded first nucleic acids and a set of longer single-stranded second nucleic acids complimentary to the first nucleic acid with a random terminal nucleotide sequence, (b) hybridizing the first nucleic acids to the second nucleic acids to form hybrids having a double-stranded portion and a single-stranded portion with the random nucleotide sequence in the single-stranded portion, (c) hybridizing a single-stranded nucleic acid target to the hybrids, (d) ligating the hybridized target to the first nucleic acid of the hybrid, (e) enzymatically extending the second nucleic acid using the target as a template, (f) isolating the extended second nucleic acid, and (g) hybridizing the first nucleic acid of step (a) with the isolated second nucleic acid to form a nucleic acid probe. It is preferred that the first nucleic acid is about 15–25 nucleotides in length, that the second nucleic acid is about 20–30 nucleotides in length, and that the double-stranded portion contain an enzyme recognition site. It is also preferred that the probe be fixed to a solid support, such as a plastic, ceramic, a metal, a resin, a gel, or a membrane. A preferred solid support is a two-dimensional or three-dimensional matrix with multiple probe binding sites, such as a hybridization chip. A further embodiment of the present invention is a diagnostic aid comprising the created nucleic acid probe and a method for using the diagnostic aid to screen a biological sample as herein described.

As an extension of this procedure, it is also possible to use the methods herein described to determine the nucleotide sequence of one or more probes which hybridize with an unknown target sequence. For example, Sanger dideoxynucleotide sequencing techniques could be used when enzymatically extending the second nucleic acid using the target as a template and labeled substrate, extended products could be resolved by polyacrylamide gel electrophoresis, and the hybridized sequences of the probes easily read off the gel. This aspect may be useful when it is cumbersome to determine the sequence of the entire target and only a smaller region of that sequence is of interest.

The following examples illustrate embodiments of the invention, but should not be viewed as limiting the scope of the invention.

EXAMPLES

Example 1

Manipulation of DNA in the solid state. Complexes between streptavidin (or avidin) and biotin represent the standard way in which much solid state DNA sequencing or other DNA manipulation is done, and one of the standard ways in which non-radioactive detection of DNA is carried out. Over the past few years streptavidin-biotin technology has expanded in several ways. Several years ago, the gene for streptavidin was cloned and sequenced (C. E. Argarana et al., Nuc. Acids Res. 14:1871, 1986). More recently, using the Studier T7 system, over-expression of the Protein in *E. coli* was achieved (T. Sano and C. R. Cantor, Proc. Natl. Acad. Sci. USA 87:142, 1990). In the last year, mutant streptavidins modified for improved solubility properties and firmer attachment to solid supports was also expressed (T. Sano and C. R. Cantor, Bio/Technology 9:1378–81, 1993). The most relevant of these is core streptavidin, (fully active protein with extraneous N- and C-terminal peptides removed) with 5 cysteine residues attached to the C-terminus. An active protein fusion of streptavidin to two IgG binding domains of staphylococcal A protein was also produced (T. Sano and C. R. Cantor, Bio/Technology 9:1378–81, 1991). This allowed biotinylated DNAs to be attached to specific Immunoglobulin G molecules without the need for any covalent chemistry, and it has led to the development of immuno-PCR, an exceedingly sensitive method for detecting antigens (T. Sano et al., Sci. 258:120–29, 1992).

A protein fusion between streptavidin and metallothionein was recently onstructed (T. Sano et al., Proc. Natl. Acad. Sci. USA, 1992). Both partners in this protein fusion are fully active and these streptavidin-biotin interactions are being used to develop new methods for purification of DNA, including triplex-mediated capture of duplex DNA on magnetic microbeads (T. Ito et al., Proc. Natl. Acad. Sci. USA 89:495–98, 1992) and affinity capture electrophoresis of DNA in agarose (T. Ito et al., G.A.T.A., 1992).

An examination of the potential advantages of stacking hybridization has been carried out by both calculations and pilot experiments. Some calculated T_m's for perfect and mismatched duplexes are shown in FIG. 1. These are based on average base compositions. The calculations were preformed using the equations given by J. G. Wetmur (Crit. Rev. in Biochem. and Mol. Biol. 26:227–59, 1991). In the case of oligonucleotide stacking, these researchers assumed that the first duplex is fully formed under the conditions where the second oligomer is being tested; in practice this may not always be the case. It will, however, be the case for the configuration shown in FIG. 1. The calculations reveal a number of interesting features about stacking hybridization. Note that the binding of a second oligomer next to a pre-formed duplex provides an extra stability equal to about two base pairs. More interesting, still, is the fact that mispairing seems to have a larger consequence on stacking hybridization than it does on ordinary hybridization. This is consistent with the very large effects seen by K. R. Khrapko

et al. (J. DNA Sequencing and Mapping 1:375–88. 1991) for certain types of mispairing. Other types of mispairing are less destabilizing, but these can be eliminated by requiring a ligation step. In standard SBH, a terminal mismatch is the least destabilizing event, and thus, leads to the greatest source of ambiguity or background. For an octanucleotide complex, an average terminal mismatch leads to a 6° C. lowering in T_m. For stacking hybridization, a terminal mismatch on the side away from the pre-existing duplex, is the least destabilizing event. For a pentamer, this leads to a drop in T_m of 10° C. These considerations indicate that the discrimination power of stacking hybridization in favor of perfect duplexes might be greater than ordinary SBH.

Example 2

Terminal sequencing by positional hybridization. The basic sequencing by hybridization scheme is depicted in FIG. 2. It is different from any other because it uses a duplex oligonucleotide array with 3'-ended single-stranded overhangs. The duplex portion of each DNA shown is constant. Only the overhangs vary, and in principle an array of 4^n probes is needed to represent all possible overhangs of length n. The advantage of such an array is that it provides enhanced sequence stringency in detecting the 5' terminal nucleotide of the target DNA because of base stacking between the preformed DNA duplex and the newly formed duplex.

One variable is the length of the single-stranded overhang. The shorter the overhang, the smaller the array of probes potentially useable. Overhangs of five and six have been successfully employed. The nature of the support surface to which the oligonucleotide is attached, the means of its attachment, and the length of the oligonucleotide duplex are also important variables. Initially one 5' end-biotinylated strand of the probe duplex is attached to a solid surface. The technology is already well developed for the attachment of nucleic acids to solid supports, such as streptavidin-coated magnetic microbeads and membranes such as the thin gel system.

Another variable is the nucleic acid capacity of the immobilized spot of probe. This determines the detection sensitivity required and is also important where unlabeled DNA may be present that could hybridize competitively with the desired labeled DNA product. As depicted in FIG. 2A, the 3' overhang of the array can detect the 3'-terminal sequence of the target DNA. These will derive from 5'-end labeled restriction fragments of known DNA sequence cut from vectors so that the target for the immobilized probe will either be at the 3' end, just internal to it, or totally internal. In some subsequent examples, it does not matter whether hybridization is absolutely specific for the 3' end.

Alternatively, positional sequencing by hybridization of the 5'-end single-stranded overhangs would be equally effective (FIG. 2B). This permits reading of the 5' terminal sequence of the target DNA. However, this approach is not as versatile because it does not allow for the use of polymerases to enhance the length and accuracy of the sequence read.

Example 3

Preparation of model arrays. Following the scheme shown in FIG. 2, in a single synthesis, all 1024 possible single-stranded probes with a constant 18 base stalk followed by a variable 5 base extension can be created. The 18 base extension is designed to contain two restriction enzyme cutting sites. Hga I generates a 5 base, 5' overhang consisting of the variable bases N_5. Not I generates a 4 base, 5' overhang at the constant end of the oligonucleotide. The synthetic 23-mer mixture will be hybridized with a complimentary 18-mer to form a duplex which can then be enzymatically extended to form all 1024, 23-mer duplexes. These can be cloned by, for example, blunt end ligation, into a plasmid which lacks Not I sites. Colonies containing the cloned 23-base insert can be selected. Each should be a clone of one unique sequence. DNA minipreps can be cut at the constant end of the stalk, filled in with biotinylated pyrimidines, then cut at the variable end of the stalk to generate the 5 base 5' overhang. The resulting nucleic acid can be fractionated by Qiagen columns (nucleic acid purification columns) to discard the high molecular weight material, and the nucleic acid probe will then be attached to a streptavidin-coated surface. This procedure could easily be automated in a Beckman Biomec or equivalent chemical robot to produce many identical arrays of probes.

The initial array contains about a thousand probes. The particular sequence at any location in the array will not be known. However, the array can be used for statistical evaluation of the signal to noise ratio and the sequence discrimination for different target molecules under different hybridization conditions. Hybridization with known nucleic acid sequences allows for the identification of particular elements of the array. A sufficient set of hybridizations would train the array for any subsequent sequencing task. Arrays are partially characterized until they have the desired properties. For example, the length of the oligonucleotide duplex, the mode of its attachment to a surface, and the hybridization conditions used, can all be varied, using the initial set of cloned DNA probes. Once the sort of array that works best is determined, a complete and fully characterized array can then be constructed by ordinary chemical synthesis.

Example 4

Preparation of specific probe arrays. The major challenge for positional SBH, is to build real arrays of probes, and test the fraction of sequences that actually perform according to expectations. Base composition and base sequence dependence on the effectiveness of hybridization is probably the greatest obstacle to successful implementation of these methods. The use of enzymatic steps, where feasible, may simplify these problems, since, after all, the enzymes do manage to work with a wide variety of DNA sequences in vivo. With positional SBH, one potential trick to compensate for some variations in stability would be to allow the adjacent duplex to vary. Thus, for an A+T rich overhang, one could use a G+C rich stacking duplex, and vice versa.

Four methods for making arrays are tested and evaluated with two major objectives. The first is to produce, rapidly and inexpensively, arrays that will test some of the principles of positional SBH. The second is to develop effective methods for the automated preparation of full arrays needed for production sequencing via positional SBH. Since the first studies indicated that a five base overhang will be sufficient, arrays may only have to have 1024 members. The cost of making all of these compounds is actually quite modest. The constant portion of the probes can be made once, and then extended in parallel, by automated DNA synthesis methods. In the simplest case, this will require the addition of only 5 bases to each of 1024 compounds, which at typical chemical costs of $2 per base will amount to a total of about $10,000.

Moderately dense arrays can be made using a typical x-y robot to spot the biotinylated compounds individually onto

a streptavidin-coated surface. Using such robots, it is possible to make arrays of 2×10^4 samples in 100 to 400 cm^2 of nominal surface. T array should preferably fit in 10 cm^2, but even if forced, for unforeseen technical reasons, to compromise on an array ten times or even 50 times less dense, it will be quite suitable for testing the principles of and many of the variations on positional SBH. Commercially available streptavidin-coated beads can be adhered, permanently to plastics like polystyrene, by exposing the plastic first to a brief treatment with an organic solvent like triethylamine. The resulting plastic surfaces have enormously high biotin binding capacity because of the very high surface area that results. This will suffice for radioactively labeled samples.

For fluorescently labeled samples, the background scattering from such a bead-impregnated sample may interfere. In this case, a streptavidin-conjugated glass or plastic surface may be utilized (commercially available from Bios Products). Surfaces are made using commercially available amine-containing surfaces and using commercially available biotin-containing N-hydroxysuccinimide esters to make stable peptide conjugates. The resulting surfaces will bind streptavidin, at one biotin binding site (or at most two, but not more because the approximate 222 symmetry of the protein would preclude this), which would leave other sites available for binding to biotinylated oligonucleotides.

In certain experiments, the need for attaching oligonucleotides to surfaces may be circumvented altogether, and oligonucleotides attached to streptavidin-coated magnetic microbeads used as already done in pilot experiments. The beads can be manipulated in microtitre plates. A magnetic separator suitable for such plates can be used including the newly available compressed plates. For example, the 18 by 24 well plates (Genetix, Ltd.; USA Scientific Plastics) would allow containment of the entire array in 3 plates; this formate is well handled by existing chemical robots. It is preferable to use the more compressed 36 by 48 well formate, so that the entire array would fit on a single plate. The advantages of this approach for all the experiments are that any potential complexities from surface effects can be avoided, and already-existing liquid handling, thermal control, and imaging methods can be used for all the experiments. Thus, this allows the characterization of many of the features of positional SBH before having to invest the time and effort in fabricating instruments, tools and chips.

Lastly, a rapid and highly efficient method to print arrays has been developed. Master arrays are made which direct the preparation of replicas, or appropriate complementary arrays. A master array is made manually (or by a very accurate robot) by sampling a set of custom DNA sequences in the desired pattern and then transferring these sequences to the replica. The master array is just a set of all 1024–4096 compounds. It is printed by multiple headed pipettes and compressed by offsetting. A potentially more elegant approach is shown in FIG. 14. A master array is made and used to transfer components of the replicas in a sequence-specific way. The sequences to be transferred are designed so that they contain the desired 5 or 6 base 5' variable overhang adjacent to a unique 15 base DNA sequence.

The master array consists of a set of streptavidin bead-impregnated plastic coated metal pins, each of which, at its tip, contains immobilized biotinylated DNA strands that consist of the variable 5 or 6 base segment plus the constant 15 base segment. Any unoccupied sites on this surface are filled with excess free biotin. To produce a replica chip, the master array is incubated with the complement of the 15 base constant sequence, 5' -labeled with biotin. Next, DNA polymerase is used to synthesize the complement of the 5 or 6 base variable sequence. Then the wet pin array is touched to the streptavidin-coated surface of the replica, held at a temperature above the T_m of the complexes on the master array. If there is insufficient liquid carryover from the pin array for efficient sample transfer, the replica array could first be coated with spaced droplets of solvent (either held in concave cavities, or delivered by a multiheaded pipettor). After the transfer, the replica chip is incubated with the complement of 15 base constant sequence to reform the double-stranded portions of the array. The basic advantage of this scheme, if it can be realized, is that the master array and transfer compounds are made only once, and then the manufacture of replica arrays should be able to proceed almost endlessly.

Example 5

DNA ligation to oligonucleotide arrays. Following the schemes shown in FIGS. 3A and 3B, *E. coli* and T4 DNA ligases can be used to covalently attach hybridized target nucleic acid to the correct immobilized oligonucleotide probe. This is a highly accurate and efficient process. Because ligase absolutely requires a correctly base paired 3' terminus, ligase will read only the 3'-terminal sequence of the target nucleic acid. After ligation, the resulting duplex will be 23 base pairs long and it will be possible to remove unhybridized, unligated target nucleic acid using fairly stringent washing conditions. Appropriately chosen positive and negative controls demonstrate the power of this scheme, such as arrays which are lacking a 5'-terminal phosphate adjacent to the 3' overhang since these probes will not ligate to the target nucleic acid.

There are a number of advantages to a ligation step. Physical specificity is supplanted by enzymatic specificity. Focusing on the 3' end of the target nucleic also minimize problems arising from stable secondary structures in the target DNA. As shown in FIG. 3B, ligation can be used to enhance the fidelity of detecting the 5'-terminal sequence of a target DNA.

DNA ligases are also used to covalently attach hybridized target DNA to the correct immobilized oligonucleotide probe. Several tests of the feasibility of the ligation scheme shown in FIG. 3. Biotinylated probes were attached to streptavidin-coated magnetic microbeads, and annealed with a shorter, complementary, constant sequence to produce duplexes with 5 or 6 base single-stranded overhangs. One set of actual sequences used is shown in Example 14. ^{32}P-end labeled targets were allowed to hybridize to the Probes. Free targets were removed by capturing the beads with a magnetic separator. DNA ligase was added and ligation was allowed to proceed at various salt concentrations. The samples were washed at room temperature, again manipulating the immobilized compounds with a magnetic separator. This should remove non-ligated material. Finally, samples were incubated at a temperature above the T_m of the duplexes, and eluted single strand was retained after the remainder of the samples were removed by magnetic separation. The eluate at this point should consist of the ligated material. The fraction of ligation was estimated as the amount of ^{32}p recovered in the high temperature wash versus the amount recovered in both the high and low temperature washes. Results obtained are shown in FIG. 13. It is apparent that salt conditions can be found where the legation proceeds efficiently with perfectly matched 5 or 6 base overhangs, but not with G-T mismatches.

The results of a more extensive set of similar experiments are shown in Tables 2–4. Table 2 looks at the effect of the

position of the mismatch and Table 3 examines the effect of base composition on the relative discrimination of perfect matches verses weakly destabilizing mismatches. These data demonstrate that: (1) effective discrimination between perfect matches and single mismatches occurs with all five base overhangs tested; (2) there is little if any effect of base composition on the amount of ligation seen or the effectiveness of match/mismatch discrimination. Thus, the serious problems of dealing with base composition effects on stability seen in ordinary SBH do not appear to be a problem for positional SBH; and (3) the worst mismatch positionis, as expected, the one distal from the phosphodiester bond formed in the ligation reaction. However, any mismatches that survive in this position will be eliminatd by a polymerase extension reaction, such as as described herein, provided that polymerase is used, like sequenase version 2, that has no 3'-endonuclease activity or terminal transferase activity; and (4) gel electrophoresis analysis has confirmed that the putative ligation products seen in these tests are indeed the actual products synthesized.

TABLE 2

Ligation Efficiency of Matched and Mismatched Duplexes in 0.2 M NaCl at 37° C. (SEQ ID NO 1) 3'-TCG AGA ACC TTG GCT-5'

		Ligation Efficiency	
	CTA CTA GGC TGC GTA GTC-5'		(SEQ ID NO 2)
5'-B-	GAT GAT CCG ACG CAT CAG AGC TC	0.170	(SEQ ID NO 3)
5'-B-	GAT GAT CCG ACG CAT CAG AGC TT	0.006	(SEQ ID NO 4)
5'-B-	GAT GAT CCG ACG CAT CAG AGC TA	0.006	(SEQ ID NO 5)
5'-B-	GAT GAT CCG ACG CAT CAG AGC CC	0.002	(SEQ ID NO 6)
5'-B-	GAT GAT CCG ACG CAT CAG AGT TC	0.004	(SEQ ID NO 7)
5'-B-	GAT GAT CCG ACG CAT CAG AAC TC	0.001	(SEQ ID NO 8)

TABLE 3

Ligation Efficiency of Matched and Mismatched Duplexes in 0.2 M NaCl at 37° C. and its Dependance on AT Content of the Overhang

Overhang Sequences		AT Content	Ligation Efficiency
Match	GGCCC	0/5	0.30
Mismatch	GGCCT		0.03
Match	AGCCC	1/5	0.36

TABLE 3-continued

Ligation Efficiency of Matched and Mismatched Duplexes in 0.2 M NaCl at 37° C. and its Dependance on AT Content of the Overhang

Overhang Sequences		AT Content	Ligation Efficiency
Mismatch	AGCTC		0.02
Match	AGCTC	2/5	0.17
Mismatch	AGCTT		0.01
Match	AGATC	3/5	0.24
Mismatch	AGATT		0.01
Match	ATATC	4/5	0.17
Mismatch	ATATT		0.01
Match	ATATT	5/5	0.31
Mismatch	ATATC		0.02

TABLE 4

Increasing Discrimination by Sequencing Extension at 37° C.

		Ligation Efficiency	Ligation Extension (cpm)	
		(percent)	(+)	(−)
	(SEQ ID NO 1) 3'-TCG AGA ACC TTG GCT-5'* CTA CTA GGC TGC GTA GTC-5'(SEQ ID NO 2)			
5'-B-	GAT GAT CCG ACG CAT CAG AGA TC (SEQ ID NO 9)	0.24	4,934	29,500
5'-B-	GAT GAT CCG ACG CAT CAG AGC TT (SEQ ID NO 10)	0.01	116	250
	Discrimination =	×24	×42	×118
	(SEQ ID NO 1) 3'-TCG AGA ACC TTG GCT-5'* CTA CTA GGC TGC GTA GTC-5'(SEQ ID NO 2)			
5'-B-	GAT GAT CCG ACG CAT CAG ATA TC (SEQ ID NO 11)	0.17	12,250	25,200

TABLE 4-continued

Increasing Discrimination by Sequencing Extension at 37° C.

	Ligation Efficiency	Ligation Extension (cpm)	
	(percent)	(+)	(−)
5'-B- GAT GAT CCG ACG CAT CAG ATA TT (SEQ ID NO 12)	0.01	240	390
Discrimination =	×17	×51	×65

"B" = Biotin
"*" = radioactive label

The discrimination for the correct sequence is not as great with an external mismatch (which would be the most difficult case to discriminate) as with an internal mismatch (Table 4). A mismatch right at the ligation point would presumably offer the highest possible discrimination. In any event, the results shown are very promising. Already there is a level of discrimination with only 5 or 6 bases of overlap that is better than the discrimination seen in conventional SBH with 8 base overlaps. Allele-specific amplification by the ligase chain reaction also appears to be quite successful (F. Baranay et al., Proc. Natl. Acad. Sci. USA 88:189–93, 1991).

Example 6

Positional sequencing by hybridization with a nested set of DNA samples. Thus far described arrays have been very inefficiently utilized because with only a single target nucleic acid, only a single probe will be detected. This clearly wastes most of the potential information intrinsically available from the array. A variation in the procedures will use the array much more efficiently. This is illustrated in FIG. 6. Here, before hybridization to the probe array, the 5'-labeled (or unlabeled) target nucleic acid is partially degraded with an enzyme such as exonuclease III. Digestion produces a large number of molecules with a range of chain lengths that share a common 5'-terminus, but have a variable 3'-terminus. This entire family of nucleic acids is then hybridized to the probe array. Assuming that the distribution of 3'-ends is sufficiently broad, the hybridization pattern should allow the sequence of the entire target to be read subject to any branch point ambiguities. If a single set of exonuclease conditions fails to provide a broad enough distribution, samples could be combined and prepared under several different conditions.

There are at least three ways to make nested DNA deletions suitable for positional SBH. The easiest, but ultimately probably the least satisfactory, is to use exonuclease, like exonuclease III, by analogy to nested deletion cloning in ordinary sequencing (S. Henikoff, Gene 28:351–58, 1984). The difficulty with these enzymes is that they may not produce an even enough yield of compounds to fully represent the sample of interest. One sees a pattern of regions in the sequence where the enzyme moves relatively rapidly, and others where it moves relatively slowly. Several commercially available enzymes can be examined by looking at the distribution of fragment lengths directly on ordinary polyacrylamide DNA sequencing gels.

The second approach to making nested samples is to use the ordinary Maxam-Gilbert sequencing chemistry. It is possible to ligate the 5'-phosphorylated fragments which result from these chemical degradations. Indeed this is the principle use for ligation-mediated genomic DNA sequencing (G. P. Pfiefer et al., Sci. 246:810–13, 1989). Asymmetric PCR or linear amplification can be used to make the complementary, ligatable, nested strands. A side benefit of this approach is that one can pre-select which base to cleave after, and this provides additional information about the DNA sequences one is working with.

The third approach to making nested samples is to use variants on plus/minus sequencing. For example, one can make a very even DNA sequencing ladder by using Sanger sequencing with a dideoxy-pppN terminator. This does not produce a ligatable end. However it can be replaced with a ligatable end, while still on the original template, by first removing the ddpppN with the 3' editing-exonuclease activity of DNA polymerase I in the absence of the one particular base at the end. Note that this accomplishes two things for the price of one. Not only does it generate a ladder with a ligatable, end, because one can pre-determine the identity of the base removed, it provides an additional nucleotide of DNA sequence information. One can use single color detection in four separate reactions, or ultimately, four color detection by mixing the results of four separate reactions prior to hybridization. If this approach is successful, it is amenable to more elaborate variations combining laddering and hybridization. Note that each of these procedures combines some of the power of ladder sequencing with the parallel processing of SBH.

In addition, there are alternative methods of preparing the desired samples, such as polymerization in the absence of limiting amounts of one of the substrate bases, such as for DNA, one of the four dNTPs. Standard Sanger or Maxam-Gilbert sequencing protocols cannot be used to generate the ladder of DNA fragments because these techniques fail to yield 3'-ligatable ends. In contrast, sequencing by the method of the present invention combines the techniques and advantages of the power of ladder sequencing with the parallel processing power of positional sequencing by hybridization.

Ligation ensures the fidelity of detection of the 3' terminal base of the target DNA. To ensure similar fidelity of detection at the 5' end of the duplex formed between the probe and the target, the probe-target duplex can be extended after ligation by one nucleotide using, for example, a labeled ddNTP (FIG. 5). This has two major advantages. First, specificity is increased because extension with the Klenow fragment of DNA polymerase requires a correctly base paired 3'-primer terminus. Second, using labeled ddNTPs one at a time, or a mixture of all four labeled with four different colors simultaneously, the identity of one additional nucleotide of the target nucleic acid can be determined as shown in FIG. 5. Thus, an array of only 1024 probes would actually have the sequencing power of an array of 4096 hexamers, in other words, a corresponding four-fold gain for any length used. In addition, polymerases work well in solid state sequencing methodologies quite analogous of the type proposed herein.

Example 7

Retaining positional information in sequencing by hybridization. Inherent in the detection of just the 3'-terminal sequence of the target nucleic acid. is the possibility of obtaining information about the distance between the sequence hybridized and a known reference point. Although that point could be arbitrary, the 5'-end of the intact target was used. The desired distance is then just the length of the DNA fragment that has hybridized to a particular probe in the array. In principle, there are two ways to determine this length. One is to length fractionate (5' labeled) DNA before or after the hybridization, ligation, and any DNA polymerase extension. Single DNA sequences could be used, but pools of many DNA targets used simultaneously or, alternatively, a double-labeled target with one color representing the 5'-end of any unique site and the other a random internal label would be more efficient. For example, incorporated into the target is a fractional amount, for example, about 1%, of biotinylated (or digoxigenin-labeled) pyrimidines, and use this later on for fluorescent detection. It has been recently shown that an internal label is effective in high sensitivity conventional ladder DNA sequencing. The ratio of the internal label to the end label is proportional to target fragment length. For any particular sample the relationship is monotonic even though it may be irregular. Thus, correct order is always obtained even if distances are occasionally distorted by extreme runs of purines of pyrimidines. If necessary, it is also possible to use two quasi-independent internal labeling schemes.

The scheme as just outlined, used with polymerase extension, might require as many as 6 different colored labels; 2 on the target (5' and internal) and four on the probe extension (four ddNTPs). However the 5' label is unnecessary, since the 3' extension provides the same information (providing that the DNA polymerase reaction is close to stoichiometric). The ddNTPs can be used one at a time if necessary. Therefore, the scheme could proceed with as little as two color detection, if necessary (FIG. 7), and three colors would certainly suffice.

A scheme complimentary to that shown in FIG. 7 would retain positional information while reading the 5'-terminal sequence of 3'-end labeled plus internally labeled target nucleic acids. Here, as in FIG. 3B, probe arrays with 5' overhangs are used, however, polymerase extension will not be possible.

Example 8

Resolution of branch point ambiguities. In current SBH, branch point ambiguities caused by sequence recurrences effectively limit the size of the target DNA to a few hundred base pairs. The positional information described in Section 6 will resolve many of these ambiguities. When a sequence recurrence occurs, if a complete DNA ladder is used as the sample, two or more targets will hybridize to the same probe. Single nucleotide additions will be informative in ¾ of the cases where two targets are ligated to the same probe; they will reveal that a given probe contains two different targets and will indicate the sequence of one base outside the recurrence. The easiest way to position the two recurrent sequences is to eliminate the longer or shorter members of the DNA ladder and hybridize remaining species to the probe array. This is a sufficiently powerful approach that it is likely to be a routine feature of positional SBH. Recurrences will be very frequent with only 5 or 6 base overhangs, but the use of segmented ladders will allow most of these to be resolved in a straightforward way. It should not be necessary to physically fractionate the DNA species of the ladder (although this could certainly be done if needed). Instead, one can cut an end-labeled ladder with a restriction nuclease. For an effective strategy seven 4-base specific enzymes should be used, singly or in combination.

Additional information is available for the recurrence of pentanucleotide sequences by the use of polymerase and single base extension as described in Example 7. In three cases out of four the single additional base will be different for the two recurrent sequences. Thus, it will be clear that a recurrence has occurred.

The real power of the positional information comes, not from its application to the recurrent sequences, but to its applications to surrounding unique sequences. Their order will be determined unequivocally, assuming even moderately accurate position information, and thus, the effect of the branch point will be eliminated. For example, 10% accuracy in intensity rations for a dual labeled 200 base pair target will provide a positional accuracy of 20 base pair. This would presumably be sufficient to resolve all but the most extraordinary recurrences.

Branch point ambiguities are caused by sequence recurrence and effectively limit the size of the target nucleic acid to a few hundred base pairs. However, positional information derived from Example 7 will resolve almost all of these ambiguities. If a sequence recurs, more than one target fragment will hybridize to, or otherwise be detected by subsequent ligation to or extension from a single immobilized probe. The apparent position of the target will be its average on the recurrent sequence. For a sequence which occurs just twice, the true location is symmetric around the apparent one. For example, the apparent position of a recurrent sequence occurring in positions 50 and 100 bases from the 5'-end of the target will be 75 bases from the end. However, when the pattern of positional sequencing by hybridization is examined, a sequence putatively located at that position will show overlap with contacts in the neighborhood of 50 bases and 100 bases from the 5'-end. This will indicate that a repeat has occurred.

Example 9

Extending the 3'-sequence of the target. Using the scheme shown in FIG. 8, it is possible to learn the identity of the base 3' to the known sequence of the target, as revealed by its hybridization position on an oligonucleotide array. For example, an array of 4^n single-stranded overhangs of the type NAGCTA 3', as shown in the Figure, are created wherein n is the number of known bases in an overhang of length n+1. The target is prepared by using a 5' label in the manner shown in FIG. 3. The Klenow fragment of DNA polymerase would then be used to add a single dpppNp as a polymerization chain terminator (or alternatively, ddpppN terminators plus ligatable ends). Before hybridization the resulting 3'-terminal phosphate would be removed by alkaline phosphatase. This would allow subsequent ligation of the target to the probe array. Either by four successive single color 5' labels, or a mixture of four different colored chains, each color corresponding to a particular chain terminator, one would be able to infer the identity of the base that had paired with the N next to the sequence AGCTA. Labeling of the 5' end minimizes interference of fluorescent base derivatives on the ligation step. Presumably, provided with a supply of dpppNp, or ribo-pppNp which can be easily prepared, the sequenase version 2 or another known polymerase will use these as a substrate. The key step in this scheme is to add a single dpppNp as a polymerization chain

terminator. Before hybridization, the resulting 3' terminal phosphate is removed by alkaline phosphatase. This allows for the subsequent ligation of the target to the probe array. Alternatively, ddpppNp terminators replaced with ligatable ends may also be used. Either by four successive single color 5' labels, or a mixture of four different colored chains, each color representing a specific chain terminator, one is able to infer the identity of the base that had paired with the N next to the sequence AGCTA. The 5' end is labeled to minimize interference of fluorescent-based derivatives with the ligation step.

Assuming that there are sufficient colors in a polychromatic detection scheme, this 3' target extension can be combined with the 3' probe extension to read n+2 bases in an array of complexity 4^n. This is potentially quite a substantial improvement. It decreases the size of the array needed by a factor of 16 without any loss in sequencing power. However, the number of colors required begins to become somewhat daunting. In principle one would want at least nine, four for each 3' extension and one general internal label for target length. However, with resonance ionization spectroscopy (RIS) detection, eight colors are available with just a single type of metal atom, and many more could be had with just two metals.

Example 10

Extending the 5' sequence of the target. In example 5, it was illustrated that by polymerase extension of the 3'-end of the probe, a single additional nucleotide on the target could be determined after ligation. That procedure used only chain terminators. Florescent labeled dNTPs that serve as substrates for DNA polymerase and other enzymes of DNA metabolism can also be made. The probe-target complex of each ligation reaction with, for example, three labeled dNTPs and a fourth unlabeled chain terminator could be extended using fluorescent labeled dNTPs. This could be repeated, successively, with each possible chain terminator. If the ratio of the intensities of the different labels can be measured fairly accurately, a considerable amount of additional sequence information will be obtained. If the absolute intensities could be measured, the power of the method appears to be very substantial since one is in essence doing a bit of four color DNA sequencing at each site on the oligonucleotide array. For example, as shown in FIG. 9, for the sequence $(Pu)_4T$, such an approach would unambiguously reveal 12 out of the 16 possible sequences and the remainder would be divided into two ambiguous pairs each. Alternatively, once the probe array has captured target DNAs, full plus-minus DNA sequencing reactions could be carried out on all targets. Single nucleotide DNA addition methods have been described that would also be suitable for such a highly parallelized implementation.

Example 11

Sample pooling in positional sequencing by hybridization. A typical 200 base pair target will detect only 196 probes on a five base 1024 probe array. This is not far from the ideal case in single, monochromatic sampling where one might like to detect half the probes each time. However, as the procedure is not restricted to single colors, the array is not necessarily this small. With an octanucleotide array, in conventional positional sequencing by hybridization or one of its herein described enhancements, the target detects only 1/32 of the immobilized probes. To increase efficiency a mixture of 16 targets can be used with two enhancements. First, intelligently constructed orthogonal pools of probes can be used for mapping by hybridization. Hybridization sequencing with these pools would be straightforward. Pools of targets, pools of probes, or pools of both can be used.

Second, in the analysis by conventional sequencing by hybridization of an array of 2×10^4 probes, divided into as few as 24 pools containing 8×10^3 probes each, there is a great deal of redundancy. Excluding branch points, 24 hybridizations could determine all the nucleic acid sequences of all the targets. However, using RIS detection there are much more than 24 colors. Therefore, all the hybridizations plus appropriate controls could be done simultaneously, provided that the density of the nucleic acid sample were high enough to keep target concentration far in excess of all the probes. A single hybridization experiment could produce 4×10^6 base pairs of sequence information. An efficient laboratory could perform 25 such hybridizations in a day, resulting in a throughput of 10^8 base-pairs of sequence per day. This is comparable to the speed of polymerization by *E. coli* DNA polymerase.

Example 12

Oligonucleotide ligation after target hybridization. Stacking hybridization without ligation has been demonstrated in a simple format. Eight-mer oligonucleotides were annealed to a target and then annealed to an adjacent 5-mer to extend the readable sequence from 8 to 13 bases. This is done with small pools of 5-mers specifically chosen to resolve ambiguities in sequence data that has already been determined by ordinary sequencing by hybridization using 8-mers alone. The method appears to work quite well, but it is cumbersome because a custom pool of 5-mers must be created to deal with each particular situation. In contrast, the approach taken herein (FIG. 9), after ligation of the target to the probe, is to ligate a mixtures of 5-mers arranged in polychromatically labeled orthogonal pools. For example, using 5-mers of the form pATGCAp or pATGCddA, only a single ligation event will occur with each probe-target complex. These would be 3' labeled to avoid interference with the ligase. Only ten pools are required for a binary sieve analysis of 5-mers. In reality it would make sense to use many more, say 16, to introduce redundancy. If only four colors are available, those would require four successive hybridizations. For example, sixteen colors would allow a single hybridization. But the result of this scheme is that one reads ten bases per site in the array, equivalent to the use of 4^{10} probes, but one only has to make 2×4^5 probes. The gain in efficiency in this scheme is a factor of 500 over conventional sequencing by hybridization.

Example 13

Synthesis of custom arrays of probes. Custom arrays of probe would be useful to detect a change in nucleic acid sequence, such as any single base change in a pre-selected large population of sequences. This is important for detecting mutations, for comparative sequencing, and for finding new, potentially rare polymorphisms. One set of target sequences can be customized to an initial general array of nucleic acid probes to turn the probe into a specific detector for any alterations of a particular sequence or series of sequences. The initial experiment is the same as outlined above in Example 4, except that the 3'-blocked 5-mers are unlabeled. After the ligation, the initial nucleic acid target strand along with its attached 18 nucleotide stalk is removed, and a new unligated 18 nucleotide stalk annealed to each element of the immobilized array (FIG. 11). The difference is that because of its history, many (ideally 50% or more), of

the elements of that array now have 10 base 3' extensions instead of 5 base extensions. These do not represent all 4^{10} possible 10-mers, but instead represent just those 10-mers which were present in the original sample. A comparison sample can now be hybridized to the new array under conditions that detect single mismatches in a decanucleotide duplex. Any samples which fail to hybridize are suspects for altered bases.

A problem in large scale diagnostic DNA sequencing is handling large numbers of samples from patients. Using the approach just outlined, a third or a fourth cycle of oligonucleotide ligation could be accomplished creating an array of 20-mers specific for the target sample. Such arrays would be capable of picking up unique segments of genomic DNA in a sequence specific fashion and detecting any differences in them in sample comparisons. Each array could be custom designed for one individual, without any DNA sequence determination and without any new oligonucleotide synthesis. Any subsequent changes in that individual's DNA such as caused by oncogenesis or environmental insult, might be easily detectable.

Example 14

Positional sequencing by hybridization. Hybridization was performed using probes with five and six base pair overhangs, including a five base pair match, a five base pair mismatch, a six base pair match, and a six base pair mismatch. These sequences are depicted in Table 5.

TABLE 5

Test Sequences:	
5 bp overlap, perfect match:	
5'-TCG AGA ACC TTG GCT*-3'	(SEQ ID NO 1)
3'-CTA CTA GGC TGC GTA GTC	(SEQ ID NO 2)
5'-biotin-GAT GAT CCG ACG CAT CAG AGC TC-3'	(SEQ ID NO 3)
5 bp overlap, mismatch at 3' end:	
5'-TCG AGA ACC TTG GCT*-3'	(SEQ ID NO 1)
3'-CTA CTA GGC TGC GTA GTC	(SEQ ID NO 2)
5'-biotin-GAT GAT CCG ACG CAT CAG AGC TT-3'	(SEQ ID NO 4)
6 bp overlap, perfect match:	
5'-TCG AGA ACC TTG GCT*-3'	(SEQ ID NO 1)
3'-CTA CTA GGC TGC GTA GTC	(SEQ ID NO 2)
5'-biotin-GAT GAT CCG ACG CAT CAG AGC TCT-3'	(SEQ ID NO 13)
6 bp overlap, mismatch four bases from 3' end:	
5'-TCG AGA ACC TTG GCT*-3'	(SEQ ID NO 1)
3'-CTA CTA GGC TGC GTA GTC	(SEQ ID NO 2)
5'-biotin-GAT GAT CCG ACG CAT CAG AGT TCT-3'	(SEQ ID NO 14)

The biotinylated double-stranded probe was prepared in TE buffer by annealing the complimentary single strands together at 68° C. for five minutes followed by slow cooling to room temperature. A five-fold excess of monodisperse, polystyrene-coated magnetic beads (Dynal) coated with streptavidin was added to the double-stranded probe, which as then incubated with agitation at room temperature for 30 minutes. After ligation, the samples were subjected to two cold (4° C.) washes followed by one hot (90° C.) wash in TE buffer (FIG. 12). The ratio of ^{32}P in the hot supernatant to the total amount of ^{32}P was determined (FIG. 13). At high NaCl concentrations, mismatched target sequences were either not annealed or were removed in the cold washes. Under the same conditions, the matched target sequences were annealed and ligated to the probe. The final hot wash removed the non-biotinylated probe oligonucleotide. This oligonucleotide contained the labeled target if the target had been ligated to the probe.

Example 15

Compensating for variations in base composition. A major problem in all suggested implementations of SBH is the rather marked dependence of T_m on base composition, and, at least in some cases, on base sequence. The use of unusual salts like tetramethyl ammonium halides or betaines (W. A. Rees et al., Biochemistry 32:137–44, 1993) offers one approach to minimizing these varieties. Alternatively, base analogs like 2,6-diamino purine and 5-bromo U can be used instead of A and T, respectively to increase the stability of A-T base paris, and derivatives like 7-deazaG can be used to decrease the stability of G-C base pairs. The initial experiments shown in Table 2 indicate that the use of enzymes will eliminate many of the complications due to base sequences. This gives the approach a very significant advantage over non-enzymatic methods which require different conditions for each nucleic acid and are highly matched to GC content.

Another method to compensate for differences in stability is to vary the base next to the stacking site. Experiments are performed to test the relative effects of all four bases in this position on overall hybridization discrimination and also on relative ligation discrimination. Base analogs such as dU (deoxyuridine) and 7-deazaG are also tested as components of the target DNA to see if these can suppress effects of secondary structure. Single-stranded binding proteins may also be helpful in this regard.

Example 16

Data measurement, processing and interpretation. Highly automated methods for raw data handling and generation of contiguous DNA sequence from the hybridization are required for analysis of the data. Two methods of data acquisition have been used in prior SBH efforts. CCD cameras with fluorescent labels and image plate analyzers with radiolabeled samples. The latter method has the advantage that there is no problem with uniform sampling of the array. However it is effectively limited to only two color analysis of DNA samples, by the use of ^{35}S and ^{32}P, differentially imaged through copper foil. In contrast, while CCD cameras are less well developed, the detection of many colors is possible by the use of appropriate exciting sources and filters. Four colors are available with conventional fluorescent DNA sequencing primers or terminators. More than four colors may be achievable if infra-red dyes are used. However, providing uniform excitation of the fluorescent array is not a trivial problem. Both detection schemes are used and the image plate analyzers are sure to work. The CCD camera approach will be necessary if some of the multicolor labeling schemes described in the proposal are ever to be realized. Label will introduced into targets by standard enzymatic methods, such as the use of 5' labeled PCR primers, for 5' labeling, internally alpha ^{32}P labeled triphosphates or fluorescent-labeled base analogs for internal labeling, and similar compounds by filling in staggered DNA ends for 3' labeling.

Both the Molecular Dynamics image plate analyzer and the Photometrics cooled CCD camera can deal with the same TIFF 8 bit data formate. Thus, software developed for either instrument can be used to handle data measured on both instruments. This will save a great deal of unnecessary duplication in data processing software. Sequence interpretation software can be developed for reading sequencing

chip data and assembling it into contiguous sequence are already underway in Moscow, at Argonne National Laboratory, and in the private sector. Such software is generally available in the interested user community. The most useful examples of this software can be customized to fit the particularly special needs of this approach including polychromatic detection, incorporation of positional information, and pooling schemes. Specific software developments for constructing and decoding the orthogonal pools of samples that may ultimately be used are being developed because these procedures are also needed for enhanced physical mapping methods.

Example 17

Generation of master beads. The general procedure for the generation of master beads is depicted in FIG. 14. Forty microliters of Dynabeads M-280 Streptavidin were washed twice with 80 µl of TE (bead concentration of 5 mg/ml). Final concentration of beads was about 5–10 pmoles of biotinylated oligo for 40 µg of beads in a total volume of 80 µl. East test oligo, in the form 5'-biotin-N_1 N_2 N_3 N_4 N_5-10 bp-3', was dissolved in TE to a concentration of 10 pmol/40 µl (250 nM). Eighty microliters of oligo were added and the mixture shaken gently for 15 minutes in a vortex at low speed.

TABLE 6

Stock solutions of MPROBEN in 1 ml Th pH 7.5			
MPROBEA	94 µg	12,200 pmol	20 µl in 1 ml
MPROBEC	121 µg	15,800 pmol	16 µl in 1 ml
MPROBEG	94 µg	12,300 pmol	20 µl in 1 ml
MPROBET	147 µg	19,200 pmol	13 µl in 1 ml
Stock solution of MCOMPBIO in 5 ml Th pH 7.5			
MCOMPBIO		464,000 pmol	5 µl in 1.85 ml

Tubes were placed in the Dynal MPC apparatus and the supernatant removed. Unbound streptavidin sites were sealed with 5 µl of 200 µM free biotin in water. Wash the beads several times with 80 µl TE. These beads can store in this state at 4° C. for several weeks.

250 nM of 5'-biotinylated 18 base nucleic acid (the complement of the constant region) served as primer for enzymatic extension of the probe region. The tube was heated to 68° C. and allowed to cool to room temperature. Beads were kept in suspension by tipping gently. Supernatant was removed and washed with 40 µl TE several times. The tube was removed from the magnet and the beads resuspended in 40 µl of TE to remove excess complement. The bead suspension was equally divided among 4 tubes and the stock tube washed with the wash divided among the tubes as well. Supernatant was removed and washed with water. Each tube contained about 2–5 pmol of DNA (28–72 ng; see Table 6).

Polymerase I extension was performed on each tube of DNA in a total of 13 µl as follows (see Table 7): NEB buffer concentration was 10 mM Tris-HCl, pH 7.5, 5 mM $MgCl_2$, 7.5 mM DTT; 33 µM d(N-N_i)TP mix; 2 µM+^{32}P dN_iTP complimentary to one of the N_i bases; and polymerase I large fragment (klenow). In the first well was added dTTP, dCTP and dGTP, to a concentration of 33 µM. ^{32}P-dATP was added to a concentration of 3 µM. dNTP stock solutions of 200 µM were pooled to lack the labelled nucleotide (i.e. Tube A contains C,G and T) adding 6.3 µl dNTP, 5 µl 200 µM dNTP, and 43 µl water. Radioactively labeled (*dNTP) stock solutions were 20 µM prepared from 2 µl [α^{32}P] dNTP, 5 µl 200 µM dNTP, and 43 µl water.

TABLE 7

TUBE #	A	C	G	T
10 × buffer	1.3 µl	1.3 µl	1.3 µl	1.3 µl
dATP	1.5 µl*	2.1 µl	2.1 µl	2.1 µl
dCTP	2.1 µl	1.5 µl*	2.1 µl	2.1 µl
dGTP	2.1 µl	2.1 µl	1.5 µl*	2.1 µl
dTTP	2.1 µl	2.1 µl	2.1 µl	1.5 µl*
Enzyme	1 µl	1 µl	1 µl	1 µl
of stock	5U	5U	5U	5U
H_2O	1.9 µl	1.9 µl	1.9 µl	1.9 µl

The tubes were incubated at 25° C. for 15 minutes. To optimize the yields of enzymatic extension, higher concentrations of dNTPs and longer reaction time may be required. The reaction was stopped by adding 4 µl of 50 mM EDTA to a final concentration of 11 µM. The supernatant was removed and the beads rinsed with 40 µl of TE buffer several times and resuspended in 35 µl of TE. The whole tube was counted and it was expected that there would be about 8% incorporation of the label added.

As a test of the synthesized oligo transfer, magnetic beads were suspended in 50 µl of 0.1M NaOH and incubated at room temperature for 10 minutes. The supernatant from each tube was removed and transfer to fresh tube. Beads were incubated a second time with 50 µl of 0.1M NaOH. As many counts seemed to remain, the first set of beads were heated to 68° C. in 50 µl NaOH which leached out a lot more counts. Each base was neutralized with 1M HCl followed by 50 µl of TE. Fresh Dynabeads were added to the melted strand and incubated at room temp for 15 minutes with gentle shaking. Supernatants were removed and saved for counting. The beads were washed several times with TE. Results are shown in Table 8.

TABLE 8

Incorporation of label (MPROBEC 5'-CATGG—)				
A	28,711/779,480			
C	35,193/574,760			
G	15,335/754,400			
T	43,048/799,440			
	Transferred	Non bound	Unmelted	Efficiency
A	9,812	2,330	10,419	43.4%
C	13,158	3,950	8,494	51.4%
G	5,621	2,672	1,924	55.0%
T	15,898	5,287	5,942	58.6%

Transferred refers to synthesized strand captured on fresh beads. Unbound refers to the synthesized strand that was not captured by the bead and unmelted refers to counts remaining on the original beads. As can be observed, between about 43% and 58% of the newly synthesized strands were successfully transferred indicating that an array of such strands could be successfully replicated.

Example 18

A procedure for making complex arrays by PCR. A slightly complex, but considerably improved scheme to test the generality of the new approach to SBH, without the need to synthesize, seprately, all 1024 five-mer probes has been developed. This procedure allows one to generate arrays with 5'- and/or 3'-overhangs and uses PCR to prepare the final probes used for hybridization which may easily be

labeled with biotin. It also builds in a way of learning part or even all of the identity of each probe sequence.

Chemical synthesis was used to make the following sequences:

(a) 5'-GTCGACAGTTGACGCTACCAYNNNNR TGGTCTAGAGCTAGC-3' (SEQ ID NO 15)

(b) 5'-CTCGAGAGTTGACGCTACCARNNNNY TGGTCTAGACCCGGG-3' (SEQ ID NO 16)

Next, enzymatic extension of the apropriate primers using a DNA polymerase in the presence of high concentrations of dNTPs was used to make the complementary duplexes. In the above sequences, N represents an equimolar mixture of all 4 bases; R is an equimolar mixture of A and G; and Y is an equimolar mixture of T and C. The underlined sequences are Bst XI and Hga I recognition sites.

-continued

5'-CTCGAG-3'	Xho I	5'-C . TCGAG-3'
3'-GAGCTC-5'	→	3'-GAGCT C-5'
5'-CCCGGG-3'	Xma I	5'-C CCGGG-3'
3'-GGGCCC-5'	→	3'-GGGCC C-5'

Those cloning sites are chosen such that, even with the degeneracy allowed by the sequences 5'-YNNNNR-3' and 5'-RNNNNY-3', these enzymes will not cleave the probe regions. For cloning, duplexes (a) were cleaved with both Sal I and Nhe I restriction enzymes (or duplexes (b) with Xho I and Xma I. The resulting digestion products were directionally cloned into an appropriate vector (e.g., plasmid, phage, etc.), suitable cells were transformed with the vector, and colonies plated. Individual clones were picked and their DNA amplified by PCR using vector

(a) 5'-GTCGACAGTTGACGCTACCAYNNNNRTGGTCTAGAGCTAGC-3' (SEQ ID NO 15)
3'-AGATCTCGATCG-5' (SEQ ID NO 17)
primer

↓

(a) 5'-GTCGACAGTTGACGCTACCAYNNNNRTGGTCTAGAGCTAGC-3' (SEQ ID NO 15)
3'-CAGCTGTCAACTGCGATGGTRNNNNYACCAGATCTCGATCG-5' (SEQ ID NO 18)

(b) 5'-CTCGAGAGTTGACGCTACCARNNNNYTGGTCTAGACCCGGG-3' (SEQ ID NO 16)
3'-AGATCTGGGCCC-5' (SEQ ID NO 19)
primer

↓

(b) 5'-CTCGAGAGTTGACGCTACCARNNNNYTGGTCTAGACCCGGG-3' (SEQ ID NO 16)
3'-GAGCTCTCAACTGCGATGGTYNNNNRACCAGATCTGGGCCC-5' (SEQ ID NO 20)

The seqences were designed with these internal Bst XI-cutting site which allows for the generation of complementary, 4 base 3'-overhanging single-strands which can be coverted to 5 base 3'-overhangs (see below) used for the type of positional SBH shown in FIG. 2A.

(SEQ ID NO 21)	5'-CCANNNNNNTGG-3'	BstX I	5'-CCANNNNN NTGG-3'
(SEQ ID NO 22)	3'-GGTNNNNNNACC-5'	→→→	3'-GGTN NNNNNACC-5'

The Hga I-cutting site overlaps with the Bst XI-cutting site and allows for the generation of 5 base 5'-overhanging single-strands. This is the structure needed for the type of postional SBH shown in FIG. 2B, and can also be used for subsequent sequencing of the overhangs by primer extension.

(SEQ ID NO 23)	5'-GACGCNNNNNNNNNN-3'	HgaI	5'-GACGCNNNNNNNNNN-3'
(SEQ ID NO 24)	3'-CTGCGNNNNNNNNNN-5'	→→→	3'-CTGCGNNNNNNNNNN-5'

The 5'- and 3'-terminal sequences of strand (a) are also recognition sites for Sal I and Nhe I, respectively; the corresponding sequence in strand (b) are recognition sites for Xho I and Xma I, respectively:

5'-GTCGAC-3'	Sal I	5'-G TCGAC-3'
3'-CAGCTG-5'	→	3'-CAGCT G-5'
5'-GCTAGC-3'	Nhe I	5'-G CTAGC-3'
3'-CGATCG-5'	→	3'-CGATC G-5'

sequences downstream and upstream from the cloned sequences as the primers. This was done to increase the length of the PCR products to ease the manipulation of these products. The probe regions from individual clones were amplified by PCR with one biotinylated primer correspond-

ing to the 5'-bases of the bottom strand. In a separate PCR, the locations of the biotins were reversed. The resulting PCR products in each case were cleaved with Bst XI, and the biotin-labeled products captured on streptavidin beads or surfaces. Note that by using PCR amplification instead of DNA purification, the need to separately purify and biotinylate each clone is also eliminated.

In parallel, all the PCR products were cleaved by Hga I which generates 5'-overhangs consisting of randomized sequences. The identity of each clone can then be determined by separate primer extensions of each of the two DNA pieces resulting from Hga I cleavage. For each pair of sequences, which derive from the same clone, the overhangs must be complementary. Therefore, sequencing just three bases on each fragment strand will given the entire structure of two probes. This plus/minus sequencing can be done in

microtire plates and is easily automated. It will fail only in the few cases were 5'-RNNNNY-3' in strand (b) contains 5'-GACGC-3', which is the recognition site for Hga I. The number of prier extension reactions required can be reduced by synthesis of more restricted pools of sequences. For example, using 4 pools where the base in one particular postion is known in advance, such as 5'-YNNANR-3'.

To make the probes needed for positional SBH (as sown in FIG. 2A), the duplex PCR products are first attached to a solid support through streptavidin. They are then cleaved with Bst XI to generate the following pairs of products:

5'-B-GTCGACAGTTGACGCTACCAYNNNN-3' (SEQ ID NO 25)

3'-CAGCTGTCAACTGCGATGGTR-5' (SEQ ID NO 26)

5'-B-GCTAGCTCTAGACCAYNNNN-3' (SEQ ID NO 27)

3'-CGATCGAGATCTGGTR-5' (SEQ ID NO 28)

5'-B-CTCGAGAGTTGACGCTACCARNNNN-3' (SEQ ID NO 29)

3'-GAGCTCTCAACTGCGATGGTY-5' (SEQ ID NO 30)

5'-B-CCCGGGTCTAGACCARNNNN-3' (SEQ ID NO 31)

3'-GGGCCCAGATCTGGTY-5' (SEQ ID NO 32)

The 5 base 3' overhangs needed for positional SBH are made by replacing the complementary (non-biotinylated) strands with constant strands which are one base shorter.

5'-B-GTCGACAGTTGACGCTACCAYNNNN-3' (SEQ ID NO 25)

3'-CAGCTGTCAACTGCGATGGT-5' (SEQ ID NO 33)

5'-B-GCTAGCTCTAGACCAYNNNN-3' (SEQ ID NO 27)

3'-CGATCGAGATCTGGT-5' (SEQ ID NO 34)

5'-B-CTCGAGAGTTGACGCTACCARNNNN-3' (SEQ ID NO 29)

3'-GAGCTCTCAACTGCGATGGT-5' (SEQ ID NO 35)

5'-B-CCCGGGTCTAGACCARNNNN-3' (SEQ ID NO 31)

3'-GGGCCCAGATCTGGT-5' (SEQ ID NO 36)

This generates the 5 base 3'-overhanging arrays amenable to extension with Sequenase version 2.0 after the ligation step shown in FIGS. 2A and B. Randomly chosen arrays of 5,120 (5× coverage) are needed to ensure that all of the sequences (>99%) are present, but this array is much larger than optimal. In practice, a library will need only provide approximately 63% of the sequences and, if necessary, can be supplemented to fill in the missing variable clones by direct synthesis.

Other embodiments and uses of the invention will be apparent to those skilled in the art from consideration of the specification and practice of the invention disclosed herein. It is intended that the specification and examples be considered exemplary only, with the true scope and spirit of the invention being indicated by the following claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i i i) NUMBER OF SEQUENCES: 48

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:1:

TCGGTTCCAA GAGCT 15

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:2:

CTGATGCGTC GGATCATC 18

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:3:

-continued

GATGATCCGA CGCATCAGAG CTC 23

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GATGATCCGA CGCATCAGAG CTT 23

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:5:

GATGATCCGA CGCATCAGAG CTA 23

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:6:

GATGATCCGA CGCATCAGAG CCC 23

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GATGATCCGA CGCATCAGAG TTC 23

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GATGATCCGA CGCATCAGAA CTC 23

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GATGATCCGA CGCATCAGAG ATC 23

-continued

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GATGATCCGA CGCATCAGAG CTT 23

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GATGATCCGA CGCATCAGAT ATC 23

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GATGATCCGA CGCATCAGAT ATT 23

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GATGATCCGA CGCATCAGAG CTCT 24

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GATGATCCGA CGCATCAGAG TTCT 24

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 41 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GTCGACAGTT GACGCTACCA YNNNNRTGGT CTAGAGCTAG C 41

-continued

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 41 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:16:

CTCGAGAGTT GACGCTACCA RNNNNYTGGT CTAGACCCGG G 41

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:17:

GCTAGCTCTA GA 12

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 41 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:18:

GCTAGCTCTA GACCAYNNNN RTGGTAGCGT CAACTGTCGA C 41

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:19:

CCCGGGTCTA GA 12

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 41 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:20:

CCCGGGTCTA GACCARNNNN YTGGTAGCGT CAACTCTCGA G 41

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:21:

CCANNNNNNT GG 12

(2) INFORMATION FOR SEQ ID NO:22:

-continued

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:22:

CCANNNNNNT GG 12

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:23:

GACGCNNNNN NNNNN 15

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:24:

NNNNNNNNNN GCGTC 15

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:25:

GTCGACAGTT GACGCTACCA YNNNN 25

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:26:

RTGGTAGCGT CAACTGTCGA C 21

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:27:

GCTAGCTCTA GACCAYNNNN 20

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 base pairs

-continued

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:28:

RTGGTCTAGA GCTAGC 16

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:29:

CTCGAGAGTT GACGCTACCA RNNNN 25

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:30:

YTGGTAGCGT CAACTCTCGA G 21

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:31:

CCCGGGTCTA GACCARNNNN 20

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:32:

YTGGTCTAGA CCCGGG 16

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:33:

TGGTAGCGTC AACTGTCGAC 20

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single

-continued

(D) TOPOLOGY: linear

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:34:

TGGTCTAGAG CTAGC 15

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:35:

TGGTAGCGTC AACTCTCGAG 20

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:36:

TGGTCTAGAC CCGGG 15

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:37:

GCAGCNNNNN NNN 13

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:38:

NNNNNNNNNN NNGCTGC 17

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:39:

GCCNNNNNGG C 11

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

-continued

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:40:

CCANNNNNNT GG 12

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:41:

GGATGNNNNN NNNN 14

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:42:

NNNNNNNNNN NNNCATCC 18

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:43:

GACGCNNNNN 10

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:44:

NNNNNNNNNN GCGTC 15

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:45:

CCANNNNNTG G 11

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:46:

-continued

GCATCNNNNN 10

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:47:

NNNNNNNNNG ATGC 14

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:48:

GGCCNNNNNG GCC 13

We claim:

1. A method for replicating an array of single-stranded probes on a solid support comprising the steps of:

a) synthesizing an array of nucleic acids each comprising a non-variant sequence of length C at a 3'-terminus and a variable sequence of length R at a 5'-terminus;

b) fixing the array to a first solid support;

c) synthesizing a set of nucleic acids each comprising a sequence complementary to the non-variant sequence;

d) hybridizing the nucleic acids of the set to the array;

e) enzymatically extending the nucleic acids of the set using the variable sequences of the array as templates;

f) denaturing the set of extended nucleic acids; and

g) fixing the denatured nucleic acids of the set to a second solid support to create the replicated array of single-stranded probes.

2. The method of claim 1 wherein the nucleic acids of the set are conjugated with biotin and the second solid support comprises streptavidin.

3. The method of claim 1 wherein the nucleic acids of the array are between about 15–30 nucleotides in length and the nucleic acids of the set are between about 10–25 nucleotides in length.

4. The method of claim 1 wherein C is between about 7–20 nucleotides and R is between about 3–5 nucleotides.

5. The method of claim 1 wherein the solid supports are selected from the group consisting of plastics, ceramics, metals, resins, gels, membranes and chips.

6. The method of claim 1 wherein the nucleic acids of the set are enzymatically extended with a DNA polymerase and one or more deoxynucleotide triphosphates.

7. The method of claim 1 wherein denaturing is performed with heat, alkali, organic solvents, binding proteins, enzymes, salts or combinations thereof.

8. The method of claim 1 further comprising the step of hybridizing the replicated array with a second set of nucleic acids complementary to the non-variant sequence of the replicated array to create a double-stranded replicated array.

9. The method of claim 1 wherein the solid supports are two-dimensional or three-dimensional matrixes.

10. The method of claim 8 wherein a double-stranded portion of the partially double-stranded replicated array comprise a restriction endonuclease site.

11. The method of claim 8 wherein the partially double-stranded replicated array have 5'- or 3'-overhangs.

12. The method of claim 8 wherein the partially double-stranded replicated array comprises about 4^R different probes.

13. The method of claim 8 wherein the partially double-stranded replicated array comprise a detectable label.

14. The method of claim 13 wherein the detectable label is selected from the group consisting of radioisotopes, stable isotopes, enzymes, fluorescent and luminescent chemicals, chromatic chemicals, metals, electric charges, and spatial chemicals.

15. A method for replicating an array of single-stranded probes on a solid support comprising the steps of:

a) fixing a first array of nucleic acids each comprising a non-variant sequence of length C at a 3'-terminus and a variable sequence of length R at a 5'-terminus to a first solid support;

b) synthesizing a first set of nucleic acids each comprising a sequence complementary to the non-variant sequence;

c) hybridizing the nucleic acids of the first set to the first array;

d) enzymatically extending the nucleic acids of the first set using the variable sequences of the first array as templates;

e) denaturing the first set of extended nucleic acids;

f) fixing the denatured nucleic acids of the first set to a second solid support to create the replicated array of single-stranded probes; and

g) hybridizing the replicated array with a second set of nucleic acids complementary to the constant sequence of the replicated array to create a partially double-stranded replicated array.

16. The method of claim 15 wherein the nucleic acids of the first set are conjugated with biotin and the second solid support comprises streptavidin.

17. The method of claim 15 wherein the nucleic acids of the first array are between about 15–30 nucleotides in length and the nucleic acids of the first set are between about 10–25 nucleotides in length.

18. The method of claim 15 wherein C is between about 7–20 nucleotides and R is between about 3–5 nucleotides.

19. The method of claim 15 wherein the partially double-stranded replicated array comprises about 4^R different probes.

20. The method of claim 15 wherein the solid supports are selected from the group consisting of plastics, ceramics, metals, resins, gels, membranes and chips.

21. The method of claim 15 wherein the solid supports are two-dimensional or three-dimensional matrixes.

22. A method for replicating an array of single-stranded probes on a solid support comprising the steps of:

a) fixing an array of nucleic acids each comprising a non-variant sequence of length C at a 3'-terminus and a variable sequence of length R at a 5'-terminus to a first solid support;

b) synthesizing a set of nucleic acids each comprising a sequence complementary to the non-variant sequence;

c) hybridizing the nucleic acids of the set to the array;

d) enzymatically extending the nucleic acids of the set using the variable sequences of the array as templates;

e) denaturing the set of extended nucleic acids; and

f) fixing the denatured nucleic acids of the set to a second solid support with a streptavidin-biotin or an avidin-biotin bond to create the replicated array of single-stranded probes.

23. The method of claim 22 wherein the nucleic acids of the array are between about 15–30 nucleotides in length and the nucleic acids of the set are between about 10–25 nucleotides in length.

24. The method of claim 22 wherein C is between about 7–20 nucleotides and R is between about 3–5 nucleotides.

25. The method of claim 22 wherein the solid supports are selected from the group consisting of plastics, ceramics, metals, resins, gels, membranes and chips.

26. The method of claim 22 wherein the nucleic acids of the set are enzymatically extended with a DNA polymerase and one or more deoxynucleotide triphosphates.

27. The method of claim 22 wherein denaturing is performed with heat, alkali, organic solvents, binding proteins, enzymes, salts or combinations thereof.

28. The method of claim 22 wherein the solid supports are two-dimensional or three-dimensional matrixes.

29. The method of claim 22 wherein the replicated array comprises about 4^R different probes.

* * * * *

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 5,795,714
DATED : 08/18/98
INVENTOR(S) : Charles R. Cantor et al.

It is certified that error appears in the above-indentified patent and that said Letters Patent is hereby corrected as shown below:

Col. 1, Line 4: Please insert --This invention was made with Government support under grant number DE-FG02-93-ER61609 awarded by the Department of Energy. The United States Government may have certain rights in this invention.--

Signed and Sealed this

First Day of June, 1999

Attest:

Q. TODD DICKINSON

Attesting Officer

Acting Commissioner of Patents and Trademarks

found to cause a partial change in the monolayer morphology resulting in some blocked channels that were not completely filled. Unfilled channels (shown as dark in Fig. 3) can be clearly distinguished from filled ones in the SFM phase image due to the difference in their viscoelasticity[30]. This material contrast has been pursued over macroscopic distances moving along the 'attolitre' liquid cavities formed by the channels (a single channel that is 1 cm in length can hold up to four attolitres).

Third, a small droplet of $FeCl_3$ solution was brought onto the structured mica surface and evaporated slowly. The $FeCl_3$ molecules condensing from the vapour phase were selectively adsorbed in the guiding channels, whereas the monolayer stripes were not coated (Fig. 4, top). The SFM images were taken at positions several centimetres away from the droplet. Channels filled with $FeCl_3$ molecules provide a contrast in magnetic force microscopy as shown in Fig. 4, bottom.

This method should not be restricted to a specific adsorbate-and-substrate pair and may be extended to other rapidly adsorbing amphiphilic molecules and polymers; however, no other distinct adsorbate-and-substrate pair has yet been investigated. The tiniest channels obtained so far with this method are about 100 nm in width corresponding to a channel density of 20,000 cm^{-1}. Such nano-channel arrays over macroscopic areas may find potential applications as chemical or biochemical reaction cavities with attolitre capacity, in analytical separation techniques, and as textured surfaces.

Received 28 June; accepted November 1999.

1. Barthlott, W. & Neinhuis, C. Purity of the sacred lotus, or escape from contaminations in biological surfaces. *Planta* 202, 1–8 (1997).
2. Wagner, T., Neinhuis, C. & Barthlott, W. Wettability and contaminability of insect wings as a function of their surface sculpture. *Acta Zool.* 77, 213–225 (1996).
3. Böltau, M., Walheim, S., Mlynek, J., Krausch, G. & Steiner, U. Surface-induced structure formation of polymer blends on patterned surfaces. *Nature* 391, 877–879 (1998).
4. Gallardo, B. S. *et al.* Electrochemical principles for active control of liquids on submillimeter scales. *Nature* 283, 57–60 (1999).
5. Delamarche, E., Bernard, A., Schmid, H., Michel, B. & Biebuyck, H. Patterned delivery of immunoglobulins to surfaces using microfluidic networks. *Science* 276, 779–781 (1997).
6. Harrison, D. J. *et al.* Micromachining a miniaturized capillary electrophoresis-based chemical analysis system on a chip. *Science* 261, 895–897 (1993).
7. Kumar, A., Biebuyk, H. A. & Whitesides, G. M. Patterning self-assembled monolayers: applications in materials science. *Langmuir* 10, 1498–1511 (1994).
8. Xia, Y. & Whitesides, G. M. Extending microcontact printing as a microlithographic technique. *Langmuir* 13, 2059–2067 (1997).
9. Evans, S. D., Flynn, T. M. & Ulman, A. Self-assembled multilayer formation on predefined templates. *Langmuir* 11, 3811–3814 (1995).
10. Abott, N. L., Folkers, J. P., Whitesides, G. M. Manipulation of the wettability of surfaces on the 0.1–1-micrometer scale through micromachining and molecular self-assembly. *Science* 257, 1380–1382 (1992).
11. Wang, R. *et al.* Light induced amphiphilic surfaces. *Nature* 388, 431–432 (1997).
12. Calvert, J. M. Lithographic patterning of self-assembled films. *J. Vac. Sci. Technol. B* 11(6), 2155–2163 (1993).
13. Gau, H., Herminghaus, S., Lenz, P. & Lipowsky, R. Liquid morphologies on structured surfaces: from microchannels to microchips. *Science* 283, 46–49 (1999).
14. Wittmann, J. C. & Smith, P. Highly oriented thin films of poly(tetrafluoroethylene) as a substrate for oriented growth of materials. *Nature* 352, 414–417 (1991).
15. Kuhn, H., Möbius, D. & Bücher, H. in *Physical Methods of Chemistry*, Vol. 1 (3b) (eds Weisenberger, A. & Rossiter, B.) 577–702 (Wiley, New York, 1972).
16. Petrov, J. G., Kuhn, H. & Möbius, D. Three-phase contact line motion in the deposition of spread monolayers. *J. Colloid Interface Sci.* 37, 66–75 (1980).
17. Riegler, H. & Spratte, K. Structural changes in lipid monolayers during the Langmuir-Blodgett transfer due to substrate / monolayer interactions. *Thin Solid Films* 210/211, 9–12 (1992).
18. Neumann, A. W. Contact angles and their temperature dependence: thermodynamic status, measurement, interpretation and application. *Adv. Colloid Interface Sci.* 59, 105–191 (1972).
19. Spratte, K., Chi, L. F. & Riegler, H. Physisorption instabilities during dynamic Langmuir wetting. *Europhys. Lett.* 25, 211–217 (1994).
20. Erikson, L. G. T., Cleasson, P. M., Ohnishi, S. & Hato, M. Stability of dimethyldioctadecylammonium bromide Langmuir-Blodgett films on mica in aqueous salt solutions—implications for surface force measurements. *Thin Solid Films* 300, 240–255 (1997).
21. Langer, J. S. Issues and opportunities in materials research. *Phys. Today* 24–31 (1992).
22. Zasadzinski, J. A. N. & Schneider, M. B. Ripple wavelength, amplitude, and configuration in lyotropic liquid crystals as a function of effective headgroup size. *J. Phys. (France)* 48, 2001–2011 (1987).
23. Maske, H. A., Havlin, S., King, P. R. & Stanley, E. Spontaneous stratification in granular mixtures. *Nature* 386, 379–382 (1997).
24. Bowden, N., Brittain, S., Evans, A. G., Hutchinson, J. W. & Whitesides, G. M. Spontaneous formation of ordered structures in thin films of metals supported on an elastomeric polymer. *Nature* 393, 146–149 (1998).
25. Weis, R. M. & McConnell, H. M. Cholesterol stabilizes the crystal-liquid interface in phospholipid monolayers. *J. Phys. Chem.* 89, 4453–4459 (1985).
26. Chunbo, Y. *et al.* Lanthanide ion induced formation of stripes domain structure in phospholipid Langmuir-Blodgett monolayers film observed by atomic force microscopy. *Surf. Sci.* 366, L729–L734 (1996).
27. Biebuyck, H. A. & Whitesides, G. M. Self-organisation of organic liquids on patterned self-assembled monolayers of alkanthiols on gold. *Langmuir* 10, 2790–2793 (1994).
28. Kim, E., Xia, Y. & Whitesides, G. M. Polymer microstructures formed by moulding in capillaries. *Nature* 376, 581–584 (1995).
29. Xia, Y. & Whitesides, G. M. Soft Lithography. *Annu. Rev. Mater. Sci.* 28, 153–184 (1998).
30. Anczykowski, B., Gotsmann, B., Fuchs, H., Cleveland, J. P. & Elings, V. B. How to measure energy dissipation in dynamic mode atomic force microscopy. *Appl. Surf. Sci.* 140, 376–382 (1999).

Acknowledgements

We thank G. Schmid for providing the Au_{55} clusters. This work was supported by the Deutsche Forschungsgemeinschaft.

Correspondence and requests for materials should be addressed to L.F.C. (e-mail:chi@nwz.uni-muenster.de).

DNA computing on surfaces

Qinghua Liu*†, Liman Wang*, Anthony G. Frutos*†, Anne E. Condon†‡, Robert M. Corn* & Lloyd M. Smith*

** Department of Chemistry, ‡ Department of Computer Science, University of Wisconsin, Madison, Wisconsin 53706, USA*

DNA computing was proposed[1] as a means of solving a class of intractable computational problems in which the computing time can grow exponentially with problem size (the 'NP-complete' or non-deterministic polynomial time complete problems). The principle of the technique has been demonstrated experimentally for a simple example of the hamiltonian path problem[2] (in this case, finding an airline flight path between several cities, such that each city is visited only once[3]). DNA computational approaches to the solution of other problems have also been investigated[4–9]. One technique[10–13] involves the immobilization and manipulation of combinatorial mixtures of DNA on a support. A set of DNA molecules encoding all candidate solutions to the computational problem of interest is synthesized and attached to the surface. Successive cycles of hybridization operations and exonuclease digestion are used to identify and eliminate those members of the set that are not solutions. Upon completion of all the multi-step cycles, the solution to the computational problem is identified using a polymerase chain reaction to amplify the remaining molecules, which are then hybridized to an addressed array. The advantages of this approach are its scalability and potential to be automated (the use of solid-phase formats simplifies the complex repetitive chemical processes, as has been demonstrated in DNA and protein synthesis[14]). Here we report the use of this method to solve a NP-complete problem. We consider a small example of the satisfiability problem (SAT)[2], in which the values of a set of boolean variables satisfying certain logical constraints are determined.

Our overall strategy for DNA computing on surfaces has been described in detail previously[10], and consists of six main steps shown diagramatically in Fig. 1. Each step in the process must be reasonably efficient for the overall process to succeed; the development of these steps has been the subject of previous work[10–12], and may be briefly summarized as follows. Oligonucleotides are synthesized individually, and pooled or arrayed on surfaces as needed. The 5′ thiol-modified oligonucleotides (S_xs) are attached to the surface in an unaddressed fashion; compared to addressed oligonucleotide

† Present addresses: Gen-Probe, Inc., San Diego, California 92121, USA (Q.L.); Corning Inc., Corning, New York 14831, USA (A.G.F.); Department of Computer Science, University of British Columbia, Vancouver, BC V6T 1Z4, Canada (A.E.C.).

arrays[15,16], this allows a greater number of different sequences to be deposited per unit area, but necessitates the use of a final 'readout' step to determine the identity of the remaining oligonucleotides once the computation has been performed[10]. Each DNA "word"[11] consists of 16 nucleotides with the structure 5′-FFFFvvvvvvvvFFFF-3′; the internal eight nucleotides ('v')are variables and encode the information (Table 1); the eight bases labelled 'F' are fixed 'word labels' that direct DNA hybridization of 16 nucleotide complements to that word in the 'mark' operation. The prototype DNA computer described here uses a single word. SAT computation is done through repeated cycles of 'mark', 'destroy' and 'unmark' operations[10] (Fig. 1). Finally, determination of the 'answer' to the computational problem of interest is accomplished in a 'readout' operation (see Methods). We note that other investigators have also reported surface-based approaches to DNA computing; indeed, Adleman's first report of DNA computing utilized support chemistry to separate DNA molecules[1]; and more recently, Morimoto *et al.*[17] and Yoshida and Suyama[18] have described solid-phase procedures for DNA computing.

The SAT problem is an NP-complete problem in boolean logic[2]. An instance of the SAT problem consists of a set of boolean logic variables separated by the logical OR operation (denoted by "∨"; $u \vee v = 0$ if and only if $u = v = 0$) within clauses, and with the clauses separated by the logical AND operation (denoted by "∧"; $u \wedge v = 1$ if and only if $u = v = 1$). The problem is to find whether there are values for the variables that simultaneously satisfy each clause in a given instance of the problem[2]. The specific SAT problem solved here is:

$$(w \vee x \vee y) \wedge (w \vee \bar{y} \vee z) \wedge (\bar{x} \vee y) \wedge (\bar{w} \vee \bar{y})$$

and employs four variables w, x, y and z ($\bar{w}$, $\bar{x}$, $\bar{y}$ and $\bar{z}$ denote the negation of the variables w, x, y and z; thus $\bar{w} = 0$ if and only if $w = 1$, and $\bar{w} = 1$ if and only if $w = 0$). Each of the four variables can be either true (1) or false (0) and thus there are total of 2^4 or 16 candidate solutions. All methods for solving the SAT problem (on any realizable computing device) that have been analysed require a number of steps that grows exponentially with the number of variables. However, algorithms have been developed for conventional computing which substantially decrease the difficulty of the problem compared to a 'brute-force' search through all possible solutions. For example, the best known algorithm for the 3-SAT problem, in which the number of variables per clause is at most 3, has expected running time bounded by a polynomial time 1.33^n for large n (ref. 19), a very substantial improvement compared to the 2^n running time characteristic of the 'brute-force' search. It has recently been shown that such algorithms may also be implemented in DNA computing approaches, including specifically the 3-SAT problem addressed here[6,18]. We show here how 3-SAT may be solved using immobilized DNA molecules and a simple 'brute-force' search strategy; the more complex operations required to execute more sophisticated algorithms are under development.

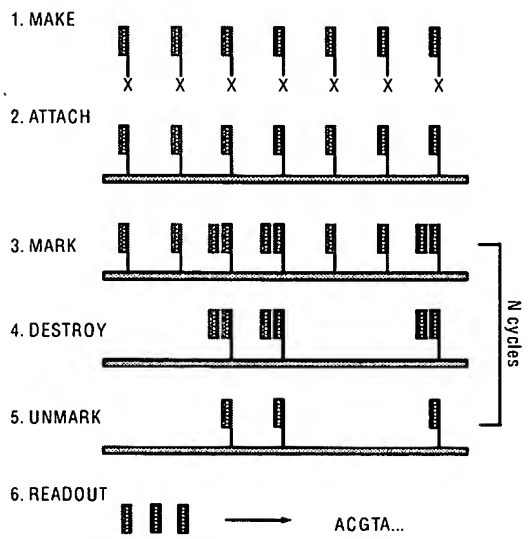


Figure 1 Overview of the surface-based approach to DNA computations. A combinatorial set of single-stranded DNA molecules representing all possible solutions to a given computational problem is synthesized ('make') and immobilized ('attach') on a surface via a reactive functional group X. In each of N successive cycles of the DNA computation, subsets of the surface-bound combinatorial mixture are tagged by hybridization to their complements in a 'mark' operation, rendering them double-stranded. After the 'mark' operation, an enzyme (for example, *Escherichia coli* exonuclease I) is added which destroys surface-bound oligonucleotides present in an unhybridized single-stranded form ('destroy'). The surface is then regenerated by removing all hybridized complements in an 'unmark' operation. Repetitive cycles of 'mark', 'destroy' and 'unmark' operations remove from the surface all strands which do not satisfy the problem. At the end of N cycles, only those strands which are solutions to the problem remain. Their identities are determined in a 'readout' operation by PCR followed by hybridization to an addressed array.

The surface-bound oligonucleotide sequences utilized were of the form 5′-HS-C_6-T_{15}GCTTvvvvvvvvTTCG-3′ (S_xs). The T_{15} sequence serves as a 'spacer' group to separate the hybridizing sequence from the support[20], the GCTT and TTCG sequences are the 'word label' used to target hybridization to a particular word[11], and the v_8 sequence is used to encode information. Table 1 shows the 16 octanucleotide sequences used, which were chosen from a previously described set of 108 possible sequences[11], along with the encoding scheme utilized to represent the possible values of the SAT variables w, x, y and z. These surface-bound oligonucleotides, representing all candidate solutions, were synthesized, pooled, and attached to a maleimide-functionalized gold surface in an unaddressed format[21,22]. Each clause of the SAT problem requires one cycle of 'mark', 'destroy' and 'unmark' (Fig. 1), and thus four cycles were employed to solve the example SAT problem above. The goal of the first computational cycle is to destroy all DNA molecules which do not satisfy the first clause ($w \vee x \vee y$). This is achieved by hybridizing to the surface those oligonucleotides that are complementary to the molecules which do satisfy the clause, and then destroying the remaining (unmarked) single-stranded molecules. Only two sequences do not satisfy this clause: namely, those for which w, x and y are set to zero (S_0 [0000] and S_1 [0001], see Table 1). Thus in cycle 1 the complements (C_xs) of the 14 other oligonucleotides ($w = 1$ (C_8, C_9, C_{10}, C_{11}, C_{12}, C_{13}, C_{14}, C_{15}); $x = 1$ (C_4, C_5, C_6, C_7, C_{12}, C_{13}, C_{14}, C_{15}); $y = 1$ (C_2, C_3, C_6, C_7, C_{10}, C_{11}, C_{14}, C_{15})) were combined and hybridized (in the 'mark' operation) to the surface; after washing, the surface was exposed to *Escherichia coli* exonuclease I to destroy the unhybridized, single-stranded molecules S_0 and S_1 (in the 'destroy' operation). The surface was regenerated

Table 1 Variable sequences and encoding scheme

Strand	Variable sequence	*wxyz*
S_0	CAACCCAA	0000
S_1	TCTCAGAG	0001
S_2	GAAGGCAT	0010
S_3	AGGAATGC	0011
S_4	ATCGAGCT	0100
S_5	TTGGACCA	0101
S_6	ACCATTGG	0110
S_7	GTTGGGTT	0111
S_8	CCAAGTTG	1000
S_9	CAGTTGAC	1001
S_{10}	TGGTTTGG	1010
S_{11}	GATCCGAT	1011
S_{12}	ATATCGCG	1100
S_{13}	GGTTCAAC	1101
S_{14}	AACCTGGT	1110
S_{15}	ACTGGTCA	1111

The 16 strands shown encode 4 bits (2^4) of information (4 variables: w, x, y, z). S_x is the 16-nucleotide DNA sequence 5′-GCTTv_8TTCG-3′, where the v_8 sequence is shown as the "Variable sequence" above; C_x is the 16-nucleotide complement of the S_x sequence.

by the 'unmark' operation to return the remaining surface-bound oligonucleotides S_2–S_{15} to single-stranded form. This process was repeated three more times for the remaining three clauses, to yield a surface containing only the solutions to the SAT problem.

In Fig. 2 we show in detail the logic of the DNA computation in each cycle, leading at the end to four types of DNA molecules remaining on the surface. It may be noted that since each clause is tested independently, in some cases complements are added corresponding to surface-bound oligonucleotides that are no longer on the surface, having been destroyed in a previous cycle. Although this entails some redundancy, the independence of clause-testing obtained in this manner increases the computational versatility and power. In the particular SAT problem solved here, each strand is destroyed only once (that is, is targeted in only one of the four computational cycles). The identity of those molecules that correspond to the solutions was determined at the end of the computation by polymerase chain reaction (PCR) amplification and hybridization to an addressed array. The results are shown in Fig. 3, presented both as a fluorescence image and in histogram form.

The four spots with high fluorescence intensity in Fig. 3 correspond to the four expected solutions to the computational problem posed. The DNA sequences identified in the 'readout' step via addressed array hybridization were: S_3, S_7, S_8 and S_9 (Fig. 3). Their variable sequences are AGGAATGC, GTTGGGTT, CCAAGTTG, and CAGTTGAC (Table 1), corresponding to the truth assignments of ($w = 0$, $x = 0$, $y = 1$, and $z = 1$), ($w = 0$, $x = 1$, $y = 1$, and $z = 1$), ($w = 1$, $x = 0$, $y = 0$, and $z = 0$), and ($w = 1$, $x = 0$, $y = 0$, and $z = 1$). For example, the assignment ($w = 0$, $x = 0$, $y = 1$, and $z = 1$) satisfies the SAT problem solved here in that $y = 1$ satisfies the first clause ($w \vee x \vee y$), $z = 1$ satisfies the second clause ($w \vee \bar{y} \vee z$), either $x = 0$ or $y = 1$ makes the third clause ($\bar{x} \vee y$) true, and $w = 0$ satisfies the last clause ($\bar{w} \vee \bar{y}$). The signal intensities for the spots corresponding to the correct solutions were 10 to 777 times greater than those corresponding to incorrect solutions, making discrimination between correct and incorrect solutions to the problem straightforward. We now consider the issues which need to be addressed for scaling of the approach to substantially larger problems.

Each step in the above-described DNA computing process is chemically complex and subject to error. For example, in the 'mark' operation, the nature of chemical reaction kinetics ensures that the hybridization-driven formation of duplex structures on the surface will be incomplete, and hence some surface-immobilized DNA molecules encoding correct answers to the clause in question will not form duplexes; these molecules will then be subject to unwanted destruction during the 'destroy' (exonuclease digestion) step, degrading signal during the computational process. Conversely, mismatch hybridization to incorrect sequences present on the surface may protect them inappropriately from destruction, leading to background signal corresponding to incorrect solutions to the problem.

In addition, the destruction of single-stranded DNA molecules on the surface by *E. coli* exonuclease I is not perfect. Approximately 94% of unmarked single-stranded immobilized DNA molecules can be removed by the exonuclease digestion reaction[11]. This means that about 6% of the unwanted (incorrect) DNA molecules will remain on the surface after a 'destroy' operation, giving rise to false positive

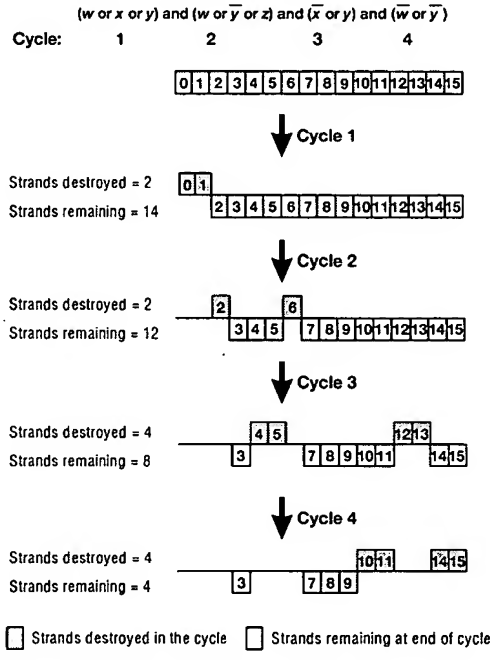


Figure 2 Four cycles of SAT computation. For each cycle of the computation, the number of strands destroyed in that cycle of the computation and the number of strands remaining on the surface at the end of the cycle are indicated on the left. The identities of the destroyed and remaining strands are shown in the progression of the computation in schematic form: in each cycle of computation one of the four clauses is tested by applying the 'mark' operation to only the oligonucleotides that satisfy that clause. For example, in cycle 1, applying the 'mark' operation to $S_2 - S_{15}$ (leaving S_0 and S_1 unmarked) protects all the surface-bound oligonucleotides for which w or x or y has a value of 1. A similar procedure is applied in each of the three subsequent cycles for clauses 2 – 4, respectively. In each cycle, strands indicated in the shaded boxes are destroyed using the single-strand specific *E. coli* exonuclease I, while strands indicated in the non-shaded boxes are protected by their complements. The strands remaining in the non-shaded boxes at the end of the process are the solutions to the computational problem.

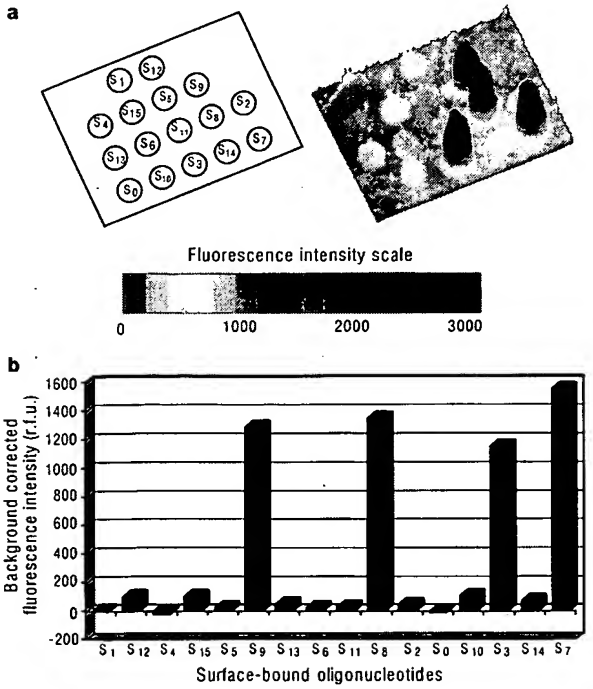


Figure 3 Three-dimensional plot and histogram of the fluorescence intensities on a 16 element addressed array used for 'readout'. **a,** Fluorescence profile (right) with the surface-bound oligonucleotide locations (left). **b,** The fluorescence intensity histogram. The three-dimensional plot in **a** was generated using NIH Image version 1.61 software (National Institutes of Health, Bethesda, Maryland; http://rsb.info.nih.gov/nih-image/download.html). r.f.u., relative fluorescence units.

signals. A more fundamental issue with the 'destroy' operation presented here is that it is incompatible with a multiple-word encoding strategy; work is in progress to develop an alternative 'destroy' operation for this purpose.

These errors in the 'mark' and 'destroy' operations also affect the 'readout' process. We have found it to be very difficult to uniformly amplify a population of DNA molecules, even ones as tightly constrained as the set employed here for DNA computing[23](data not shown). Small amounts of unwanted oligonucleotides, if amplified preferentially compared to desired species present initially at higher concentration, can yield aberrant results. Although the GC content was held constant for the set of DNA molecules used in this work, variations in the position of GC pairs within the sequence can affect the formation of secondary structures that influence the efficiency of PCR amplification. Betaine is known to preferentially destabilize GC base pairs[24]. The use of betaine here resulted in a more uniform amplification of the mixture of DNA templates[25]. Nonetheless, substantial optimization and experimentation with the PCR conditions were required to obtain the results shown here, and it is likely that this would become increasingly difficult with larger problems. This is fundamentally a reflection of the intrinsically nonlinear nature of the PCR amplification process[26]; an alternative method of amplification which is linear in nature is under investigation[27].

This intrinsically error-prone nature of the DNA computing process, however, does not rule out its operational practicality. The fundamental operations of conventional computing are also prone to error at the device level, and there is a rich body of knowledge on algorithmic solutions to these issues which could be adapted to the DNA computing process[28]. Methods for handling errors in DNA computations have been reported[29,30]. Experimental determination of error rates in surface-based DNA computing could provide the information necessary to develop error models and algorithmic methods to control errors in computations.

The surface approach described here was designed to permit scale-up to larger problems. For example, the use of six tandem words on the surface (for a total oligomer length of $16 \times 6 = 96$ nucleotides, within the range of current oligonucleotide synthesis capabilities), and 64 different octanucleotide sequences[11] to encode information in each of the words (6 bits per word), corresponds to a search space containing approximately 6.9×10^{10} candidate solutions (2^{36}; that is, a 36-bit DNA computer).

The number of 'mark' and 'destroy' operations required to solve the problem grows polynomially with the number of variables[13], in contrast to the search space which grows exponentially. Thus a 36-bit DNA computer would not look a great deal different from the 4-bit computer described here. The biggest practical issue (apart from those mentioned above) is the synthesis and handling of large numbers of oligonucleotides (1,536 oligonucleotides would be needed to carry out the 'mark' and readout' operations for a 6-word, 6 bits per word, DNA computer of this type; see Methods). Instrumentation to synthesize and handle such large numbers of oligonucleotides is available (ABI 3948 Nucleic Acid Synthesis and Purification System (Perkin Elmer); also see refs 15 and 16). Specialized problems might require far fewer oligonucleotides than this (see Methods).

Methods

PCR amplification for 'readout'.

56-mer PCR templates were synthesized with the following structure:

5'-tatttttgagcagtggctccCGAAvvvvvvvvAAGCtagctatctacaagattcag-3'

where the lower-case sequences are the two primers, the upper-case sequences are fixed word labels, and the eight bases labelled 'v' are complements of the variable regions shown in Table 1.

For PCR amplification and addressed array hybridization 'readout', the template(s) used in PCR amplification were collected from the computation chip (that is, the non-addressed array surface which had undergone repetitive computation operations) as follows: upon completion of each computation cycle (including 'mark', 'destroy', 'unmark'), the coverslip (computation chip) was exposed to a mixture of all 16 complementary DNA strands (each containing two 20-mer primer sequences as indicated above) for 1.5 h for hybridization. The coverslip was then washed; all hybridized complements were melted off and collected by assembling the coverslip in a GeneAmp *In Situ* PCR System 1000 (Perkin Elmer) containing 100 μl of de-ionized water, and heating at 95 °C for 10 min. The collected template(s) were then amplified using biotinylated and fluorescently tagged PCR primers as follows:

5'-biotin-ctgaatcttgtagatagcta-3' 5'-FAM-tatttttgagcagtggctcc-3'

The amplification reaction (100 μl) contained 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 3 mM $MgCl_2$, 0.2 mM each dATP, dCTP, dTTP and dGTP, 1 μM of each primer, 1 M betaine, and 5 units of AmpliTaq DNA polymerase (Perkin Elmer). In addition, 10 fmol of a 56-mer oligonucleotide identical in structure to that shown above, but containing in the variable region the non-complementary 8-mer sequence GAGACTCT, was included in the reactions. The function of this was to help equalize the amount of material that could be amplified in the various PCR reactions performed. This was found to reduce artefacts due to amplification of trace amounts of templates present in the reactions, a source of background in the 'readout' operation. PCR was performed in an PTC-200 Peltier Thermal Cycler (MJ Research, Inc.) using 25 cycles. The PCR mixture was first held at 95 °C for 1 min, followed by cycles of 95 °C for 30 s, 40 °C for 20 s and 60 °C for 1 min.

Hybridization to addressed array for 'readout'.

The DNA addressed array was prepared by methods similar to those described in ref. 31. Briefly, a gold-coated coverslip was immersed in a 0.5 mM ethanolic solution of *n*-octadecylmercaptan overnight, followed by photopatterning of the surface using an aluminium mask consisting of 1.5-mm-diameter holes with 3 mm centre-to-centre spacing. The coverslip was then immersed in 1 mM ethanolic 11-mercaptoundecanoic acid for 1 h, followed by electrostatic adsorption of a poly-L-lysine monolayer, which was maleimide-activated as described previously[21].

PCR amplification of the products of the DNA computing process yields double-stranded 56-mers with one strand biotinylated, and the other strand fluorescein-labelled. Before addressed array readout, these double-stranded PCR products were strand-separated using streptavidin-coupled magnetic beads (Dynabeads M-280, Dynal Prod. No. 112.05, 112.06 Protocol, 1997, Dynal Inc.). After strand separation, the fluorescently labelled PCR products were resuspended in hybridization buffer (HB; 2XSSPE/0.2%SDS), applied to the 16-spot addressed array and allowed to hybridize for 1.5 h at room temperature. The addressed array coverslip was washed twice in HB, 5 min per wash, followed by immersion in a beaker of HB at 37 °C for 20 min. The coverslip was then scanned using a Molecular Dynamics FluorImager 575.

Number of oligonucleotides required for scale-up.

The oligonucleotides employed for a general purpose multiple word DNA computer of the type employed here are of two types: (1) multiple word combinatorial mixtures, thiol-modified to permit surface attachment when preparing the unaddressed surface on which the computation is performed, and (2) single word individual oligonucleotides, either thiol-modified for surface attachment when preparing addressed 'readout' arrays (S_xs), unmodified for use in the 'mark' operation (C_xs), or with PCR-primer sequences appended for use in the 'readout' operation (C_xs with appended PCR-primer sequences). Thus a total of four sets of oligonucleotides must be synthesized. Although (as described in the text) the number of distinct oligonucleotides present in the combinatorial mixtures can be very large, a much smaller set of distinct oligonucleotide synthesis procedures is required to prepare them using split-and-pool methodology[32,33]; this is given by the product of the number of different oligonucleotide sequences per word and the number of words. In the 36-bit example provided above, this corresponds to $64 \times 6 = 384$ oligonucleotide syntheses. The number of individual oligonucleotides required for the three sets of S_xs and C_xs is given by the same product, for a total of $384 \times 4 = 1,536$; when solving a 3-SAT, in any given cycle (corresponding to a single clause) the number of C_xs required for the 'mark' operation is at most $(64 / 2) \times 3 = 96$.

Received 7 July; accepted 20 October 1999.

1. Adleman, L. M. Molecular computation of solutions to combinatorial problems. *Science* 266, 1021–1024 (1994).
2. Garey, M. R. & Johnson, D. S. *Computers and Intractability: a Guide to the Theory of NP-completeness* (Freeman, New York, 1979).
3. Adleman, L. M. Computing with DNA. *Sci. Am.* 279, 54–61 (1998).
4. Lipton, R. J. DNA solution of hard computational problems. *Science* 268, 542–545 (1995).
5. Guarnieri, F., Fliss, M. & Bancroft, C. Making DNA add. *Science* 273, 220–223 (1996).
6. Ogihara, M. *Breadth First Search 3SAT Algorithms for DNA Computers* (Technical report TR 629, Department of Computer Science, University of Rochester, Rochester, New York, 1996).
7. Ouyang, Q., Kaplan, P. D., Liu, S. & Libchaber, A. DNA solution of the maximal clique problem. *Science* 278, 446–449 (1997).
8. Seeman, N. C. *et al.* New motifs in DNA nanotechnology. *Nanotechnology.* 9, 257–273 (1998).
9. Winfree, E., Liu, F. R., Wenzler, L. A. & Seeman, N. C. Design and self-assembly of two-dimensional DNA crystals. *Nature* 394, 539–544 (1998).
10. Smith, L. M. *et al.* A surface-based approach to DNA computation. *J. Comput. Biol.* 5, 255–267 (1998).
11. Frutos, A. G. *et al.* Demonstration of a word design strategy for DNA computing on surfaces. *Nucl. Acid. Res.* 25, 4748–4757 (1997).
12. Frutos, A. G., Smith, L. M. & Corn, R. M. Enzymatic ligation reactions of DNA "words" on surfaces for DNA computing. *J. Am. Chem. Soc.* 120, 10277–10282 (1998).

13. Cai, W. *et al.* in *Proc. 1st Annu. Int. Conf. on Computational Molecular Biology (RECOMB97)* 67–74 (Association for Computing Machinery, New York, 1997).
14. Smith, L. M. Automated synthesis and sequence analysis of biological macromolecules. *Anal. Chem.* 60, 381A–390A (1988).
15. Fodor, S. P. A. *et al.* Light-directed, spatially addressable parallel chemical synthesis. *Science* 251, 767–773 (1991).
16. Chee, M. *et al.* Accessing genetic information with high-density DNA arrays. *Science* 274, 610–614 (1996).
17. Morimoto, N., Arita, M. & Suyama, A. in *Proc. DIMACS: DNA Based Computers (III)* (eds Rubin, H. & Wood, D. H.) 83–92 (American Mathematical Society, Providence, 1997).
18. Yoshida, H. & Suyama, A. in *Preliminary Proc. DIMACS: DNA Based Computers (V)* 9–20 (American Mathematical Society, Providence, 1999).
19. Schöning, U. in *Proc. 40th Annu. IEEE Conf. of Foundations of Computer Science (FOCS)* 410–414 (IEEE Computer Society, Los Alamitos, California, 1999).
20. Guo, Z., Guilfoyle, R. A., Thiel, A. J., Wang, R. & Smith, L. M. Direct fluorescence analysis of genetic polymorphisms by hybridization with oligonucleotide arrays on glass supports. *Nucl. Acid. Res.* 22, 5456–5465 (1994).
21. Jordan, C. E., Frutos, A. G., Thiel, A. J. & Corn, R. M. Surface plasmon resonance imaging measurements of DNA hybridization adsorption and streptavidin/DNA multilayer formation at chemically modified gold surfaces. *Anal. Chem.* 69, 4939–4947 (1997).
22. Bain, C. D. *et al.* Formation of monolayer films by the spontaneous assembly of organic thiols from solution onto gold. *J. Am. Chem. Soc.* 111, 321–335 (1989).
23. Baskaran, N. *et al.* Uniform amplification of a mixture of deoxyribonucleic acids with varying GC content. *Genome Res.* 6, 633–638 (1996).
24. Rees, W. A., Yager, T. D., Korte, J. & von Hippel, P. H. Betaine can eliminate the base pair composition dependence of DNA melting. *Biochemistry* 32, 137–144 (1993).
25. Voss, K. O., Pieter Roos, K., Nonay, R. L. & Dovichi, N. J. Combating PCR bias in bisulfate-based cytosine methylation analysis. Betaine-modified cytosine deamination PCR. *Anal. Chem.* 70, 3818–3823 (1998).
26. Farrell, R. E. DNA amplification. *Immunol. Invest.* 26, 3–7 (1997).
27. Lyamichev, V. *et al.* Polymorphism identification and quantitative detection of genomic DNA by invasive cleavage of oligonucleotide probes. *Nature Biotechnol.* 17, 292–296 (1999).
28. Pippenger, N. Developments in the synthesis of reliable organisms from unreliable components. *Proc. Symp. Pure Math.* 50, 311–324 (1990).
29. Boneh, D., Dunworth, C., Lipton, R. J. & Sgall, J. DNA Based Computers II (eds Landweber, L. F. & Baum, E. B.) 163–170 (DIMACS Series in Discrete Mathematics and Theoretical Computer Science, Vol. 44, American Mathematical Society, Providence, 1999).
30. Karp, R. M., Kenyon, C. & Waarts, O. Error-resilient DNA computation. *Random Struct. Algor.* 15, 450–466 (1999).
31. Gillmor, S. D., Thiel, A. J., Smith, L. M. & Lagally, M. G. Hydrophilic/hydrophobic patterned surfaces as templates for DNA arrays. *Langmuir* (submitted).
32. Gallop, M. A., Barrett, R. W., Dower, W. J., Fodor, S. P. A. & Gordon, E. M. Applications of combinatorial technologies to drug discovery. 1. Background and peptide combinatorial libraries. *J. Med. Chem.* 37, 1233–1251 (1994).
33. Gordon, E. M., Barrett, R. W., Dower, W. J., Fodor, S. P. A. & Gallop, M. A. Applications of combinatorial technologies to drug discovery. 2. Combinatorial organic synthesis, library screening strategies, and future directions. *J. Med. Chem.* 37, 1385–1401 (1994).

Acknowledgements

We thank S. Gillmor and J. Brockman for help with the preparation of the photopatterned read-out arrays, and M. Lagally for discussions. This work was supported by the Defense Advanced Research Projects Agency (DARPA) and the National Science Foundation.

Correspondence and requests for materials should be addressed to L.M.S. (e-mail: smith@chem.wisc.edu)

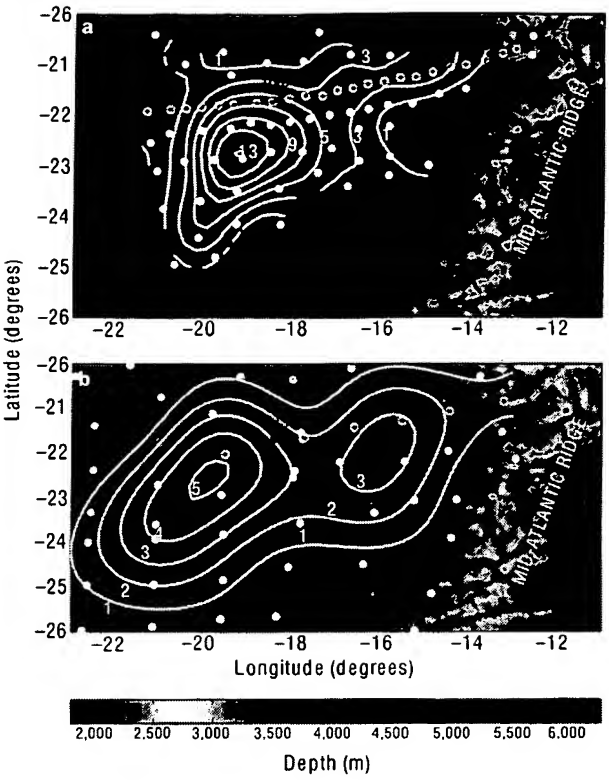


Figure 1 Tracer distribution. **a**, 14 months after release; **b**, 26 months after release. The red bars labelled 'INJ' mark the release site of the tracer. The contours denote the column integral of SF_6 (in nmol m^{-2}) and colours denote bottom depth. The tracer mapping procedure did not take the bathymetry into account, and hence the meridional distribution in the east in **a** appears broader than the valleys where the stations are located and which actually hold most of the tracer. The bathymetry is from Smith and Sandwell[16]. The stations are shown as white dots; those with stars are used for the sections in Fig. 2. Southerly latitude and westerly longitude are shown negative.

Evidence for enhanced mixing over rough topography in the abyssal ocean

J. R. Ledwell, E. T. Montgomery, K. L. Polzin, L. C. St. Laurent, R. W. Schmitt & J. M. Toole

Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

The overturning circulation of the ocean plays an important role in modulating the Earth's climate. But whereas the mechanisms for the vertical transport of water into the deep ocean—deep water formation at high latitudes—and horizontal transport in ocean currents have been largely identified, it is not clear how the compensating vertical transport of water from the depths to the surface is accomplished. Turbulent mixing across surfaces of constant density is the only viable mechanism for reducing the density of the water and enabling it to rise. However, measurements of the internal wave field, the main source of energy for mixing, and of turbulent dissipation rates, have typically implied diffusivities across surfaces of equal density of only ~0.1 $cm^2 s^{-1}$, too small to account for the return flow. Here we report measurements of tracer dispersion and turbulent energy dissipation in the Brazil basin that reveal diffusivities of 2–4 $cm^2 s^{-1}$ at a depth of 500 m above abyssal hills on the flank of the Mid-Atlantic Ridge, and approximately 10 $cm^2 s^{-1}$ nearer the bottom. This amount of mixing, probably driven by breaking internal waves that are generated by tidal currents flowing over the rough bathymetry, may be large enough to close the buoyancy budget for the Brazil basin and suggests a mechanism for closing the global overturning circulation.

Our study was conducted in the abyssal Brazil basin where deep upwelling can be inferred from measurements of net inflow of dense water[4,5]. In 1996 we surveyed turbulent microstructure and internal wave fine-structure across the basin, and released 110 kg of sulphur hexafluoride above one of the zonal valleys on the western flank of the Mid-Atlantic Ridge[6] (Fig. 1). The microstructure data showed diapycnal mixing to be very low over the smooth parts of the Brazil

US006534271B2

(12) **United States Patent**
Fürste et al.

(10) **Patent No.: US 6,534,271 B2**
(45) **Date of Patent: Mar. 18, 2003**

(54) **CLONING AND COPYING ON SURFACES**

(75) Inventors: **Jens Peter Fürste**, Berlin (DE); **Sven Klussmann**, Berlin (DE); **Thomas Klein**, Berlin (DE); **Günter Von Kiedrowski**, Bochum (DE)

(73) Assignee: **Noxxon Pharma AG**, Berlin (DE)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **09/866,513**

(22) Filed: **May 25, 2001**

(65) **Prior Publication Data**

US 2002/0022275 A1 Feb. 21, 2002

Related U.S. Application Data

(63) Continuation-in-part of application No. PCT/DE99/03856, filed on Nov. 26, 1999.

(30) **Foreign Application Priority Data**

Nov. 27, 1998 (DE) 198 54 946

(51) **Int. Cl.**[7] **C12Q 1/68**; C12P 19/34; G01N 33/53

(52) **U.S. Cl.** **435/6**; 435/7.1; 435/91.2

(58) **Field of Search** 435/6, 91.1, 91.2, 435/7.1; 436/94

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,458,066	A	7/1984	Caruthers et al.	
4,683,195	A	7/1987	Mullis et al.	
4,683,202	A	7/1987	Mullis	
5,424,186	A	6/1995	Fodor et al.	
5,445,934	A	8/1995	Fodor et al.	
5,489,678	A	2/1996	Fodor et al.	
5,510,270	A	4/1996	Fodor et al.	
5,641,658	A	6/1997	Adams et al.	
5,700,637	A	12/1997	Southern	
5,795,714	A	8/1998	Cantor et al.	
6,103,463	A *	8/2000	Chetverin et al.	435/6
6,280,950	B1 *	8/2001	Lipshutz et al.	435/6

FOREIGN PATENT DOCUMENTS

EP	0374665	6/1990
WO	9317126	9/1993
WO	9429484	12/1994
WO	9601836	1/1996
WO	9604404	2/1996

(List continued on next page.)

OTHER PUBLICATIONS

Systematic Functional Analysis of the Yeast Genome, Stephen G. Oliver, Michael K. Winson, Douglas B. Kell and Frank Baganz, T/B Tech, Sep. 1998, vol. 16, pp. 373–378.
"RNA-catalysed carbon—carbon bond formation," Theodore M. Tarasow, Sandra L. Tarasow, and Bruce E. Eaton, Nature, vol. 389, pp. 54–57.

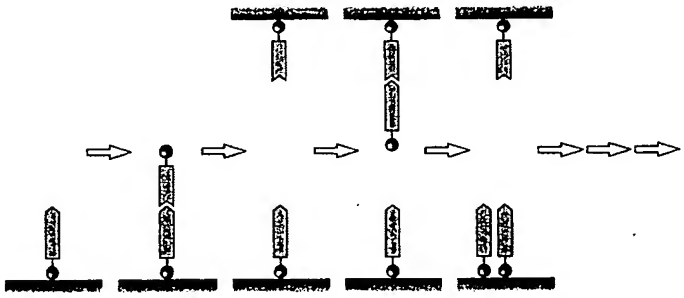
(List continued on next page.)

Primary Examiner—Kenneth R. Horlick
(74) *Attorney, Agent, or Firm*—Nils H. Ljungman & Associates

(57) **ABSTRACT**

The present invention provides for amplification methods for cloning and copying genetic material on surfaces as well as copying biological material insofar as, in a broader sense, it can be classified as a ligand-receptor system. The invention therefore relates in particular to a method for propagating ligands and receptors on at least two surfaces, comprising (a) immobilizing a first ligand on a first surface of a substantially solid phase; (b) adding a solution of receptors and binding complementary receptors to the first ligand; (c) transferring the receptor to a second surface and immobilizing the receptor at that location; (d) attaching an additional ligand to the immobilized receptor; and (e) transferring the additional ligand to the first surface and immobilizing it at that location, wherein the steps set forth above may be repeated,multiple times.

22 Claims, 16 Drawing Sheets



FOREIGN PATENT DOCUMENTS

WO	9706468		2/1997
WO	9707243		2/1997
WO	9809735		3/1998
WO	9814610		4/1998
WO	9820019		5/1998
WO	99/19341	*	4/1999
WO	00/27521	*	5/2000

OTHER PUBLICATIONS

"Macromolecular matchmaking: advances in two–hybrid and related technologies," Robert M. Frederickson, Analytical Biotechnology, pp. 90–96.

"Automated DNA Sequencing and Anaylsis," Mark D. Adams, ed.

"Cloning Differentially Expressed mRNAs," Jackson S. Wan, et al., Nature Biotechnology, vol. 14, pp. 1685–1691.

"Present State and Frontiers in Biosensories," F.W. Scheller, E. Schubert, and J. Ecdrowitz, pp. 1–9.

"Why pharmacogenomics? Why now?, " David Housman and Fred D.Ledley, Nature Biotechnology, vol. 16, pp. 492–493.

"Light–Directed Spatially Addressable Parallel Chemical Synthesis," Stephen P.A. Foder, et al., Research Article, pp. 767–773.

"Surface–promoted replication and exponential amplification of DNA analogues," A. Luther, et al., Nature, vol. 396, pp. 245–248.

* cited by examiner

Fig. 1A

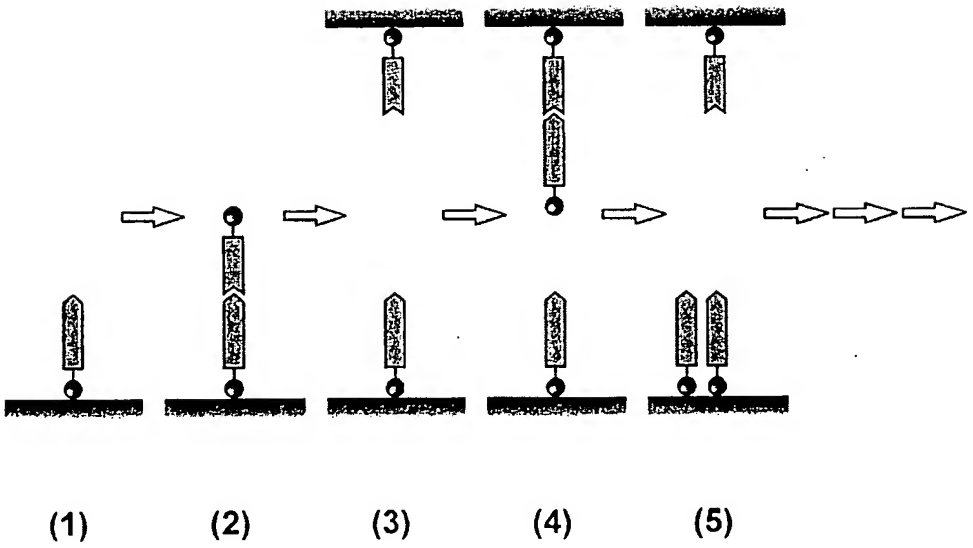


Fig. 1B

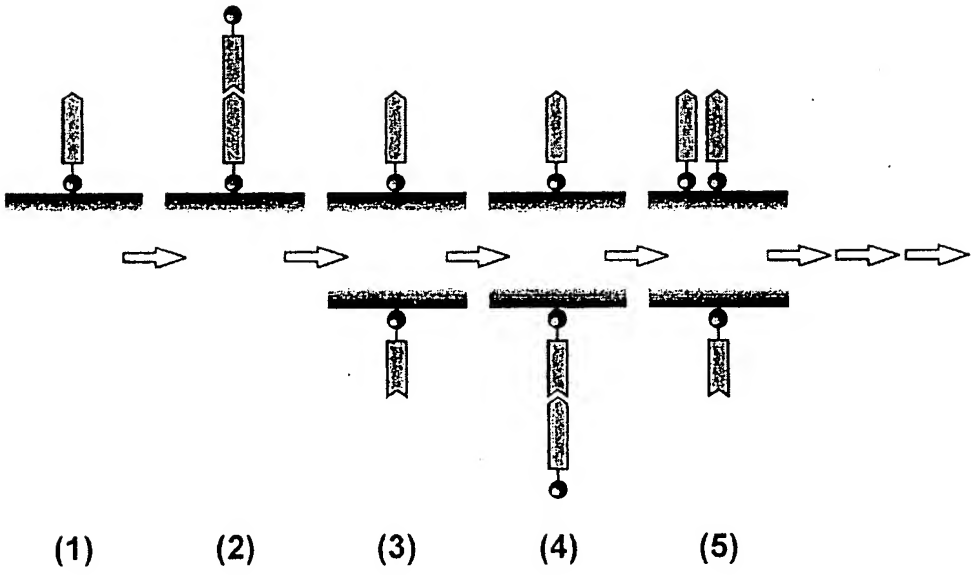


Fig. 2

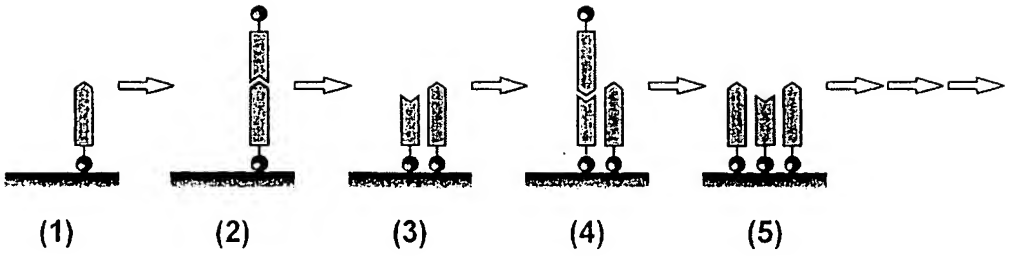
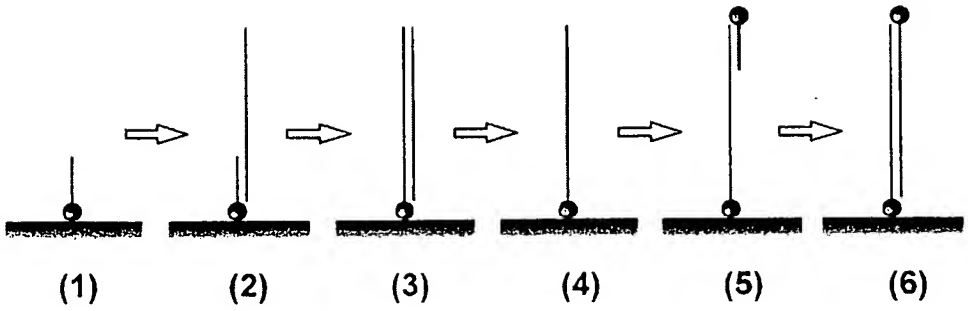


Fig. 3A



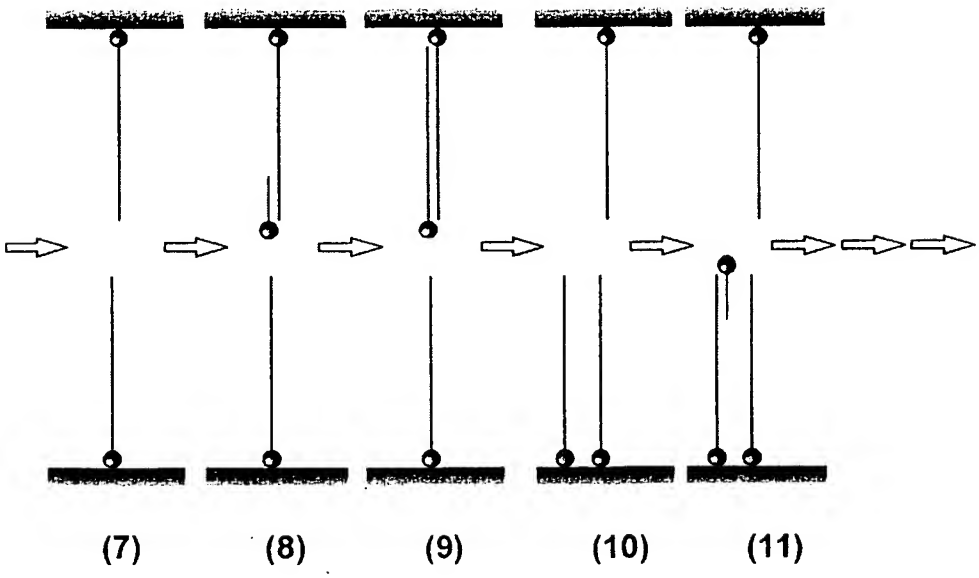


Fig. 3B

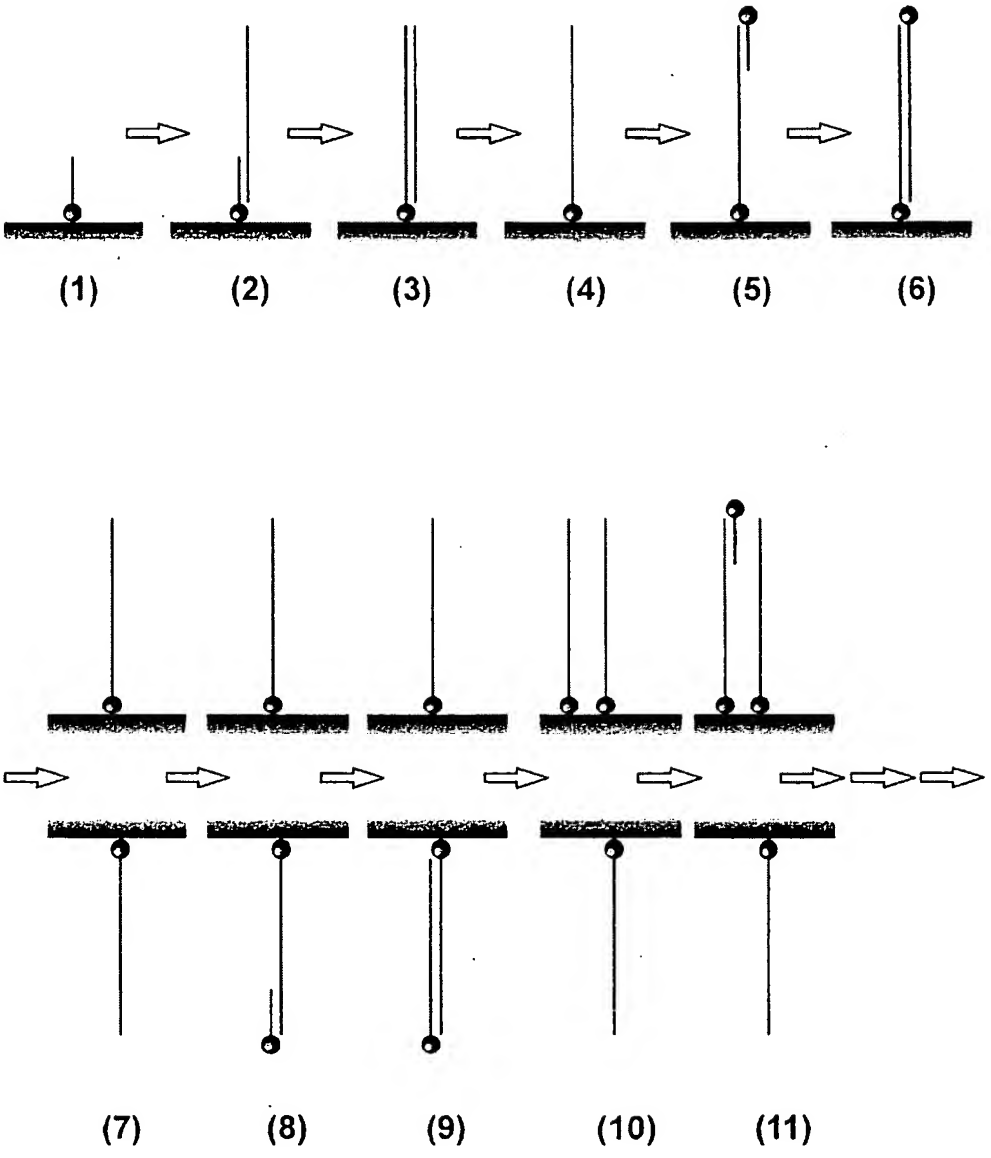
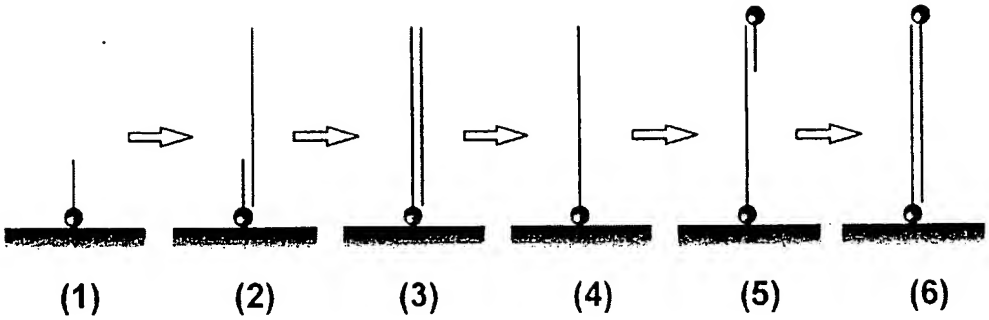


Fig. 4



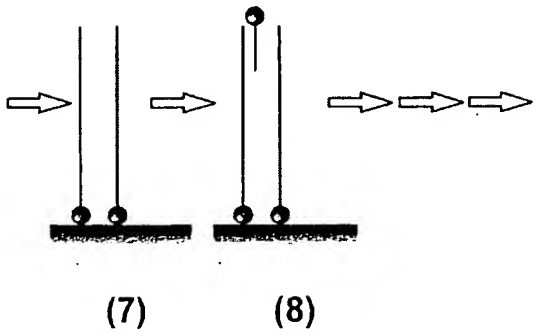


Fig. 5

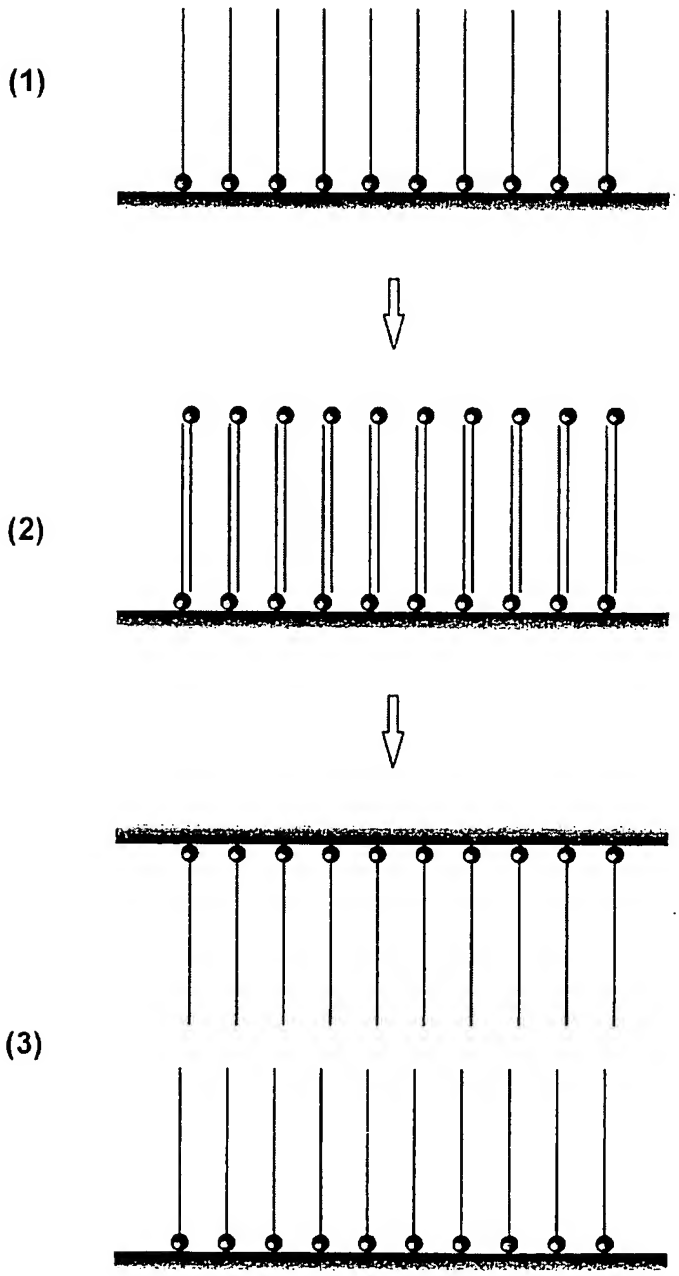


Fig. 6

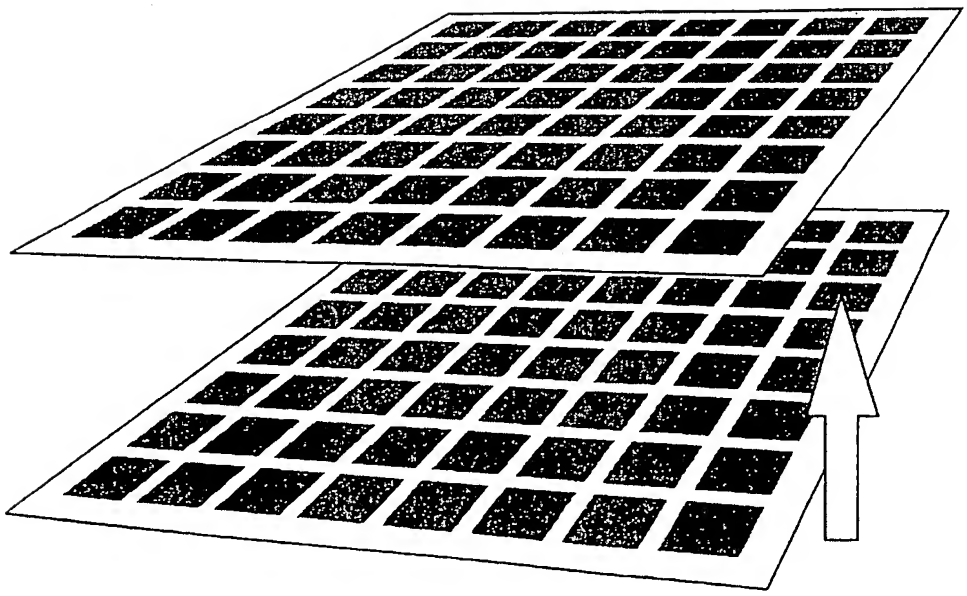
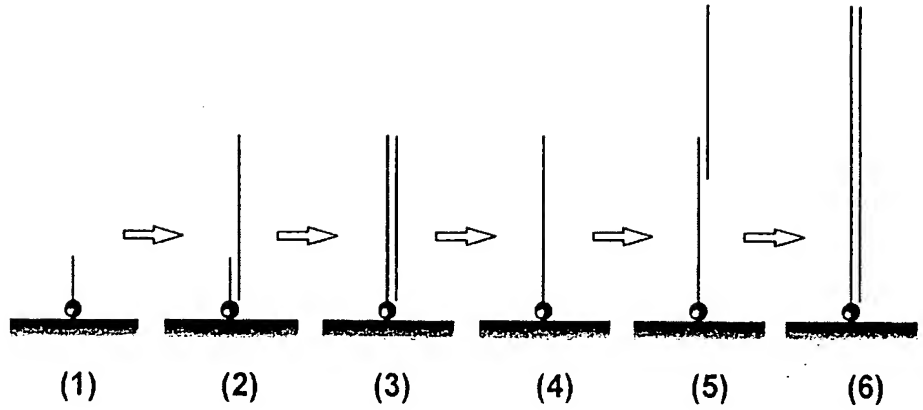


Fig. 7



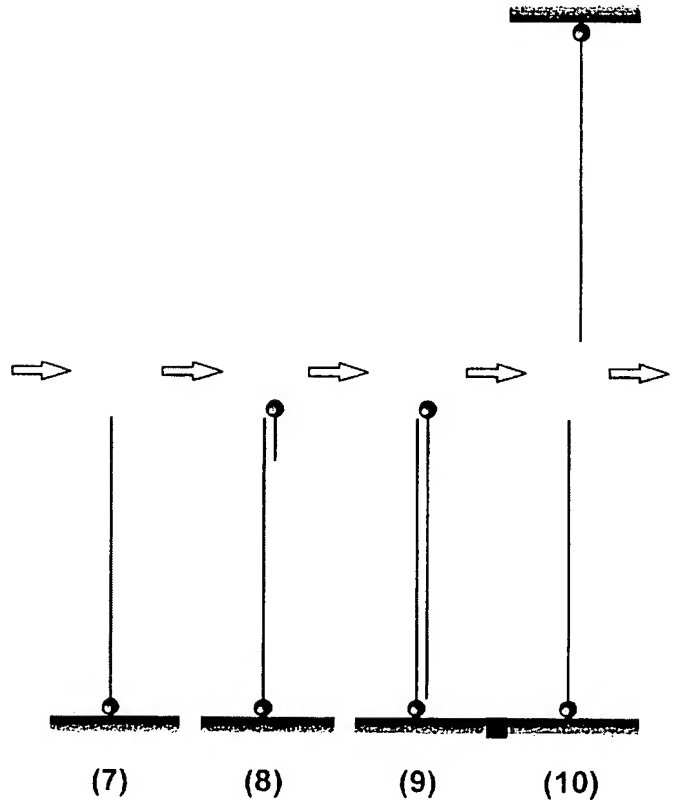


Fig. 8

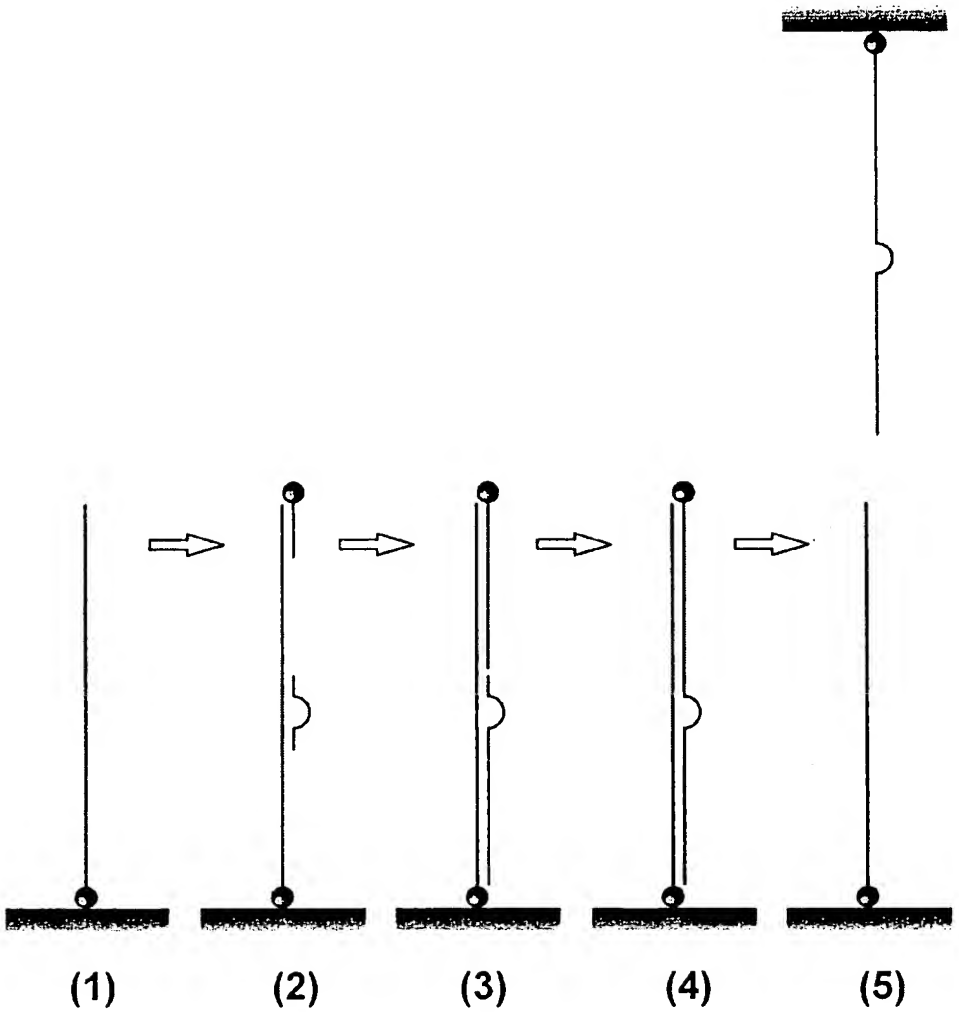


Fig. 9

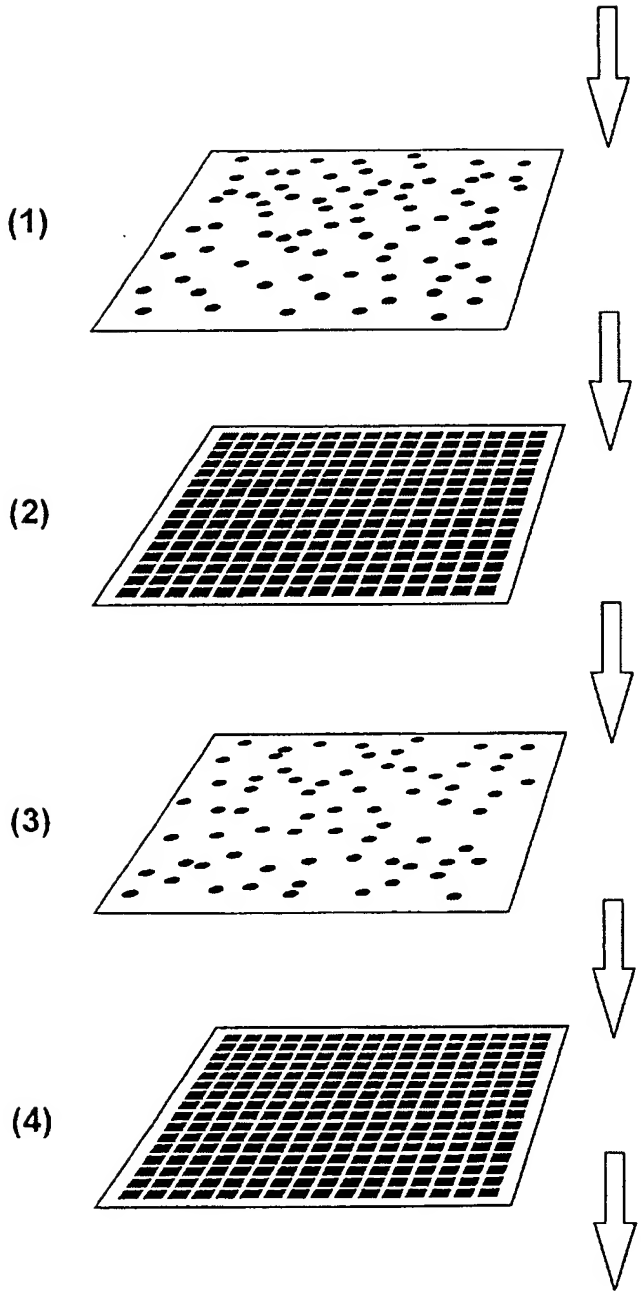


Fig. 10

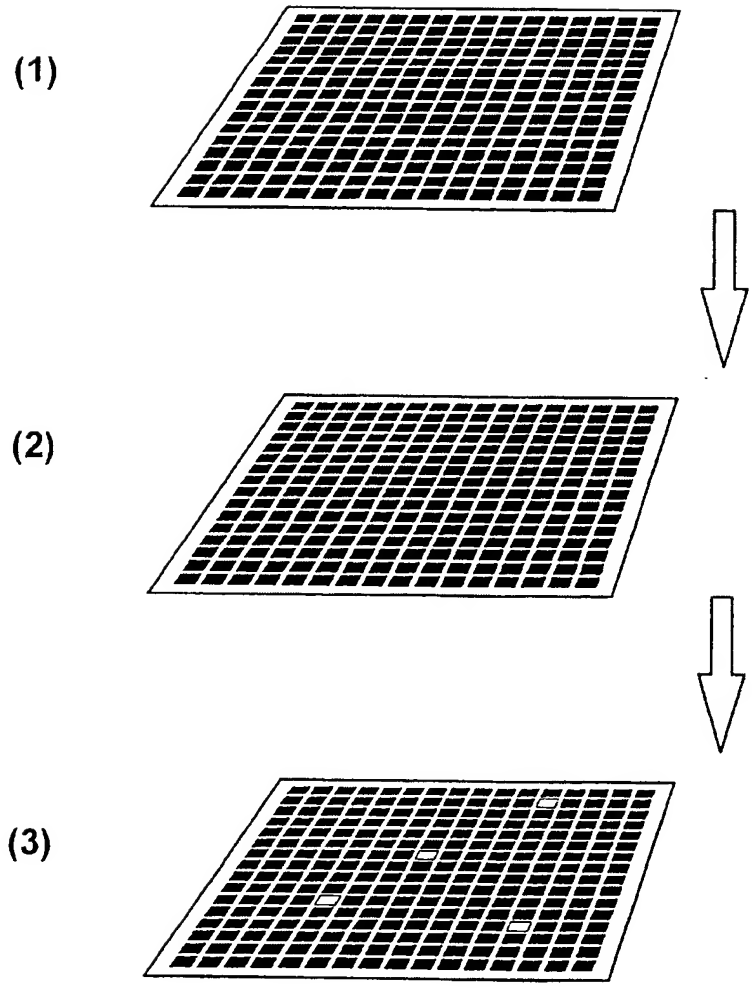


Fig. 11

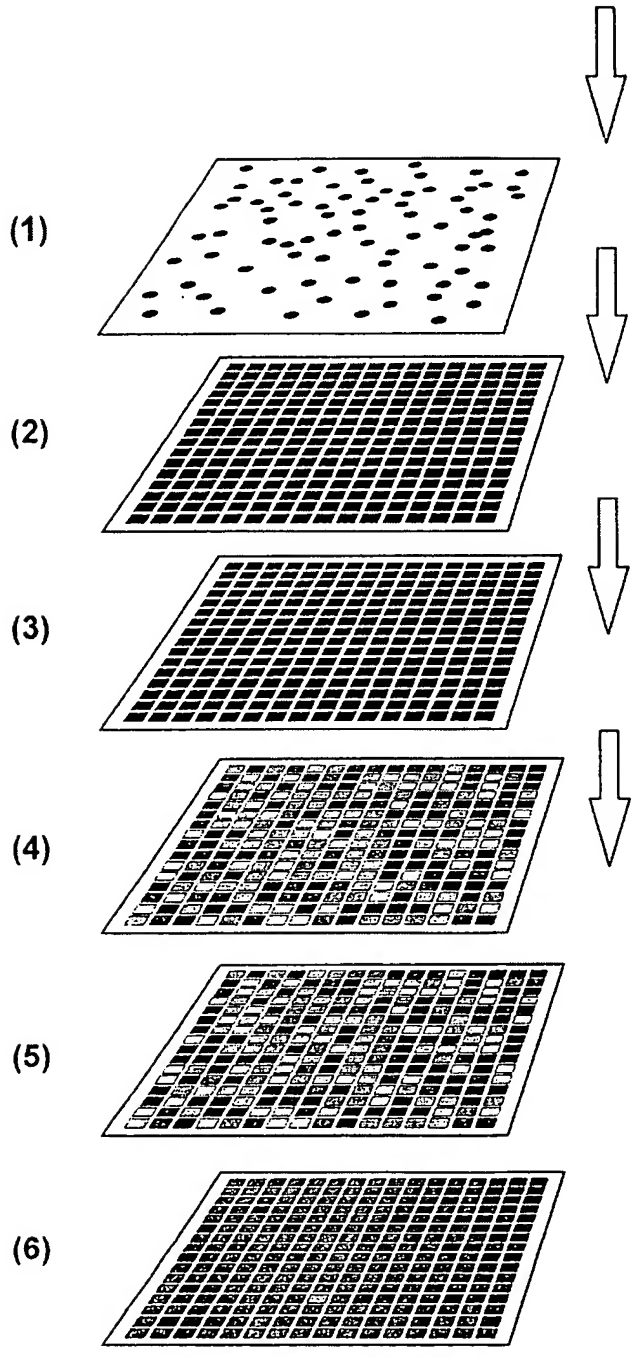


Fig. 12

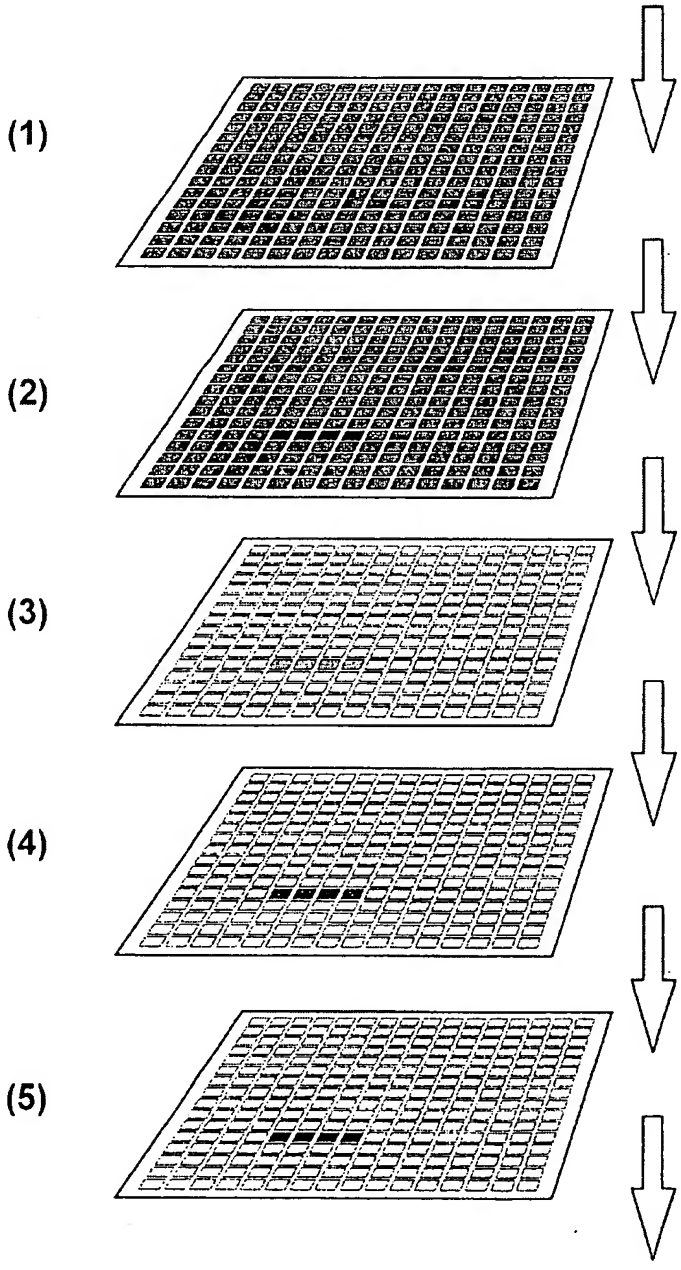


Fig. 13

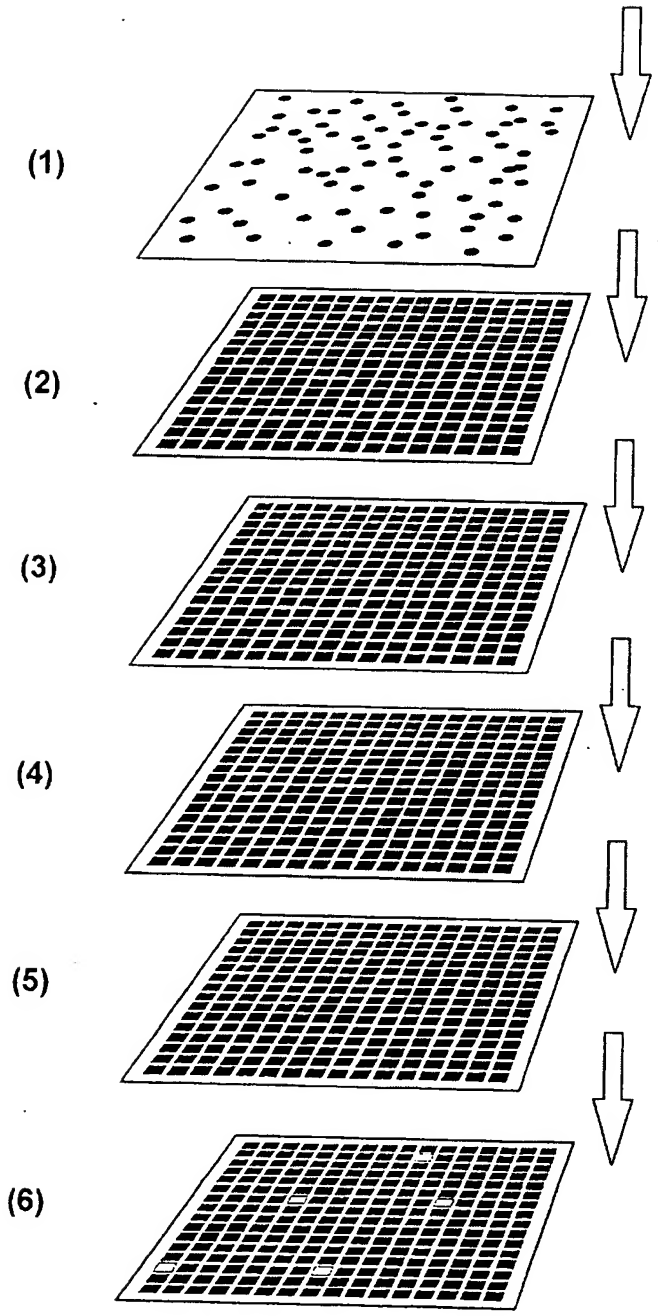
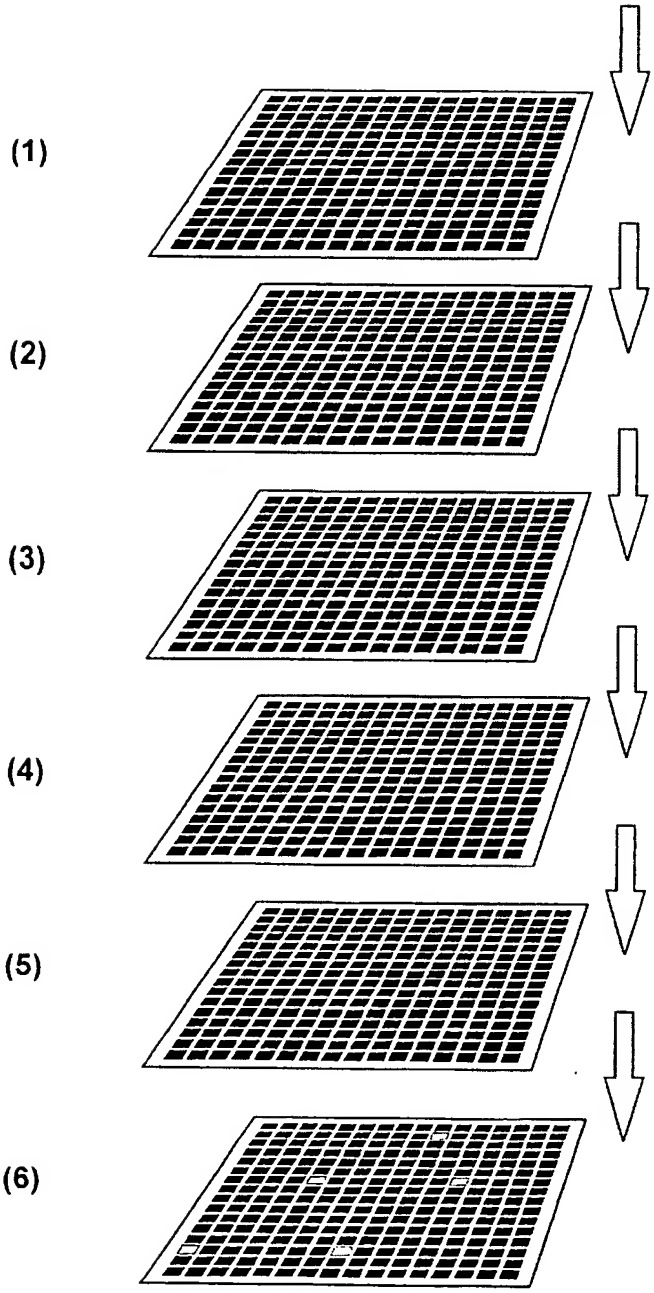


Fig. 14



CLONING AND COPYING ON SURFACES

CONTINUING APPLICATION DATA

This application is a Continuation-In-Part application of International Patent Application No. PCT/DE99/03856, filed on Nov. 26, 1999, which claims priority from Federal Republic of Germany Patent Application No. DE 198 54 946.6, filed on Nov. 27, 1998. International Application No. PCT/DE99/03856 was pending as of the filing date of the above-cited application. The United States was an elected state in International Application No. PCT/DE99/03856.

BACKGROUND OF THE INVENTION

1. Field of the Invention

This invention relates to a method for cloning and copying genetic material on surfaces as well as copying biological material insofar as it can be classified in a broader sense in a ligand-receptor system.

2. Background Information

Methods of exponentially amplifying molecular matrices are already known through the work of G. von Kiedrowski et al. (Nature 1998, Vol. 346, 245–248; Federal Republic of Germany Patent No. 198 48 403). The amplification cycles are characterized by:

- binding of molecular matrices to the surface of a solid phase by means of a reversible linker on the matrix;
- addition of matrix fragments, with one of the fragments displays a linker unit, which may, if necessary, be protected;
- synthesizing copies of the matrix;
- removal of superfluous matrix fragments and ancillary reaction substances;
- detachment of the copies from the matrix; and
- application of synthesized matrix copies to free binding sites on the solid phase.

This represents an iterative, progressive amplification method, allowing an exponential increase in the amount of molecular matrices available, thus enabling a significant process of evolution to take place. To achieve this the process makes use of the surface of a solid carrier. Chemical attachment to immobilized matrices enables copies to be synthesized from precursor matrices, which are then released to become new matrices. This process can be repeated any number of times.

In addition, so-called "bridge" amplification technology is described in U.S. Pat. No. 5,641,658. This is an amplification model based on conventional PCR methods, but which is intended to achieve localized amplification. Bridge amplification technology has many uses, especially in analytical methods that can also be carried out with the commonly used PCR. The bridge technology facilitates the separation and detection stages of the amplified products. The characteristics of this technology are that it combines amplification, selection and detection in a single process. Advanced state of the art systems can be found on the homepage of MOSAIC Technologies, Inc. (USA), the company marketing bridge amplification technology (www.mostek.com).

The bridge system describes a method for amplifying nucleic acids on a solid phase, with both amplification primers being bound covalently to a single solid phase through their 5' ends. Consequently this represents a further development of the well-known polymerase chain reaction, known as PCR for short. This takes place in a solid phase PCR instead of in a solution. The particular advantage of this method is its ability to amplify and analyze many different

genetic elements simultaneously using a single sample. The applications for bridge amplification technology include genetic expression, genome research, clinical diagnostics and the examination of biological fluids, e.g. blood. A higher rate of amplification is achieved by eliminating ineffective primer artifacts (such as primer dimers). This enables simple, sensitive and cost-effective DNA detection methods to be developed, for example using fluorescence. Because bridge amplification technology ensures that all amplification products remain bound to the solid phase, contamination through prolongation remains low, and this in turn enhances the diagnostic value of the method compared with the usual PCR.

Whereas the method proposed by G. von Kiedrowski et al. referred to above demonstrates the benefits of the solid phase amplification of entire populations, the method described in U.S. Pat. No. 5,641,658 offers the advantage of amplifying a single matrix on a solid phase. However, the drawback of this bridge amplification method is linked with the problem of product inhibition, i.e. a newly produced copy may occur not only with the adjacent immobilized primer, but also with the original matrix strand, which is also adjacent. Another disadvantage is the lower linear limitation needed to achieve bridging as a double strand. Furthermore, there is no separation between the strands, with the result that, for diagnostic purposes, the hybridization signals are weakened due to hybridization with complementary strands.

Federal Republic of Germany Patent No. 694 09 646 T2 describes a method for amplifying a nucleic acid, in which the one primer is bound to a solid phase and the second primer to a particle that reacts with a magnetic field. These primers are incorporated in target nucleic acid sequences. Following an extension stage the nucleic acid strands are separated by the application of an electric current. The magnetic primer can be particle-bonded, existing as a form of solid phase. The avidin/biotin system is suitable for binding the primer to the solid phases. This method is also suitable for cloning.

U.S. Pat. No. 5,795,714 describes a method which, in one form, uses an array of oligonucleotides, which are connected to the surface of the solid phases by means of the reciprocal reaction between biotin and avidin. The method described consists of the hybridization of complementary strands, primer extension reactions, the hybridization of a second biotinylized primer to the primer extension products, and the extension of the second primer. Mention is made of the blotting of copies on a second surface which is coated with avidin.

The disadvantage of the methods referred to above is that multiple replication in the sense of exponential propagation is not possible, and that the translocation of the copies by means of an electromagnetic field cannot be achieved without loss of site information. Based on this state of the art, and avoiding the shortcomings referred to above, it is therefore the task of this invention to provide a method of cloning and copying onto surfaces which permits the propagation of biological material while retaining site information.

SUMMARY OF THE INVENTION

The present invention overcomes the problems and disadvantages of current amplifications methods and enables biological systems including, nucleic acids, ligands and receptors, to be propagated and separated from one another by means of an electric field for immobilization and fixation on one or more solid surfaces, while retaining site information.

The present invention considers a biological system to be basically an interaction between nucleic acids of any kind

and/or with peptides/proteins/polymerases/enzymes (DNA/RNA/PNA/pRNA/2'–5' nucleotides and RNA/DNA mirror mers (see PCT/EP97/04726)), in exactly the same way as antigen/antibody complexes or, in general terms, ligand/receptor systems.

For a basic understanding of the invention it is necessary to appreciate that complementary nucleic acids themselves represent nothing more than a special form of a complementary ligand/receptor system in a traditional sense. For purposes of the present application, the term "ligand" will refer to one molecule of the biological binding pair and the term "receptor" will refer to the opposite molecule of the biological binding pair. Two complementary strands of nucleic acid are biological binding pairs. One of the strands is designated the ligand and the other strand is designated the receptor. Biological binding pairs may also comprise antigen and antibodies, drugs and drug receptor sites and enzymes and enzyme substrates.

The invention utilizes the fact that, because of their charge, nucleic acids and many other biologically relevant molecules can be moved within an electrical field when such a field is applied. In the present case a stationary, bound molecule is separated from a corresponding molecule in this way, in that the non-stationary, bound molecule is either separated from the other by synthesization of that molecule or following an "identification reaction", with the aid of an electrical field. Because the molecules tend to migrate along the line of the electrical field, they retain site information while migrating. This is also what mainly distinguishes the method referred to in this invention from the previously known state of the art, because in Federal Republic of Germany Patent No. 694 09 646 T2, for example, site information is not retained, since the electromagnetic field described in that case does not serve to retain the site information.

For all of the alternative processes shown hereinbelow it is possible to alter, reduce and/or increase the (geometrical) scale of the transfer, while retaining site information, within the respective transfer stage of the respective process sequence. This may be practical, for example, if the geometry of the arrays from which or to which the transfer is being made is not identical with that of the initial array or target array.

The present invention in one aspect relates broadly to a method for propagating ligands and receptors on at least two surfaces, comprising:

(a) immobilizing a first ligand on a first surface of a substantially solid phase;

(b) adding a solution of receptors and binding complementary receptors to the first ligand;

(c) transferring the receptor to a second surface and immobilizing the receptor at that location;

(d) attaching an additional ligand to the immobilized receptor; and

(e) transferring the additional ligand to the first surface and immobilizing it at that location, wherein the steps set forth above may be repeated multiple times.

Preferably, the surface in step (c) is a second surface which is spatially separated from the first. The additional ligand is transferred to the second surface by the application of an electrical field, and preferably the electrical field is applied between the first and second surface.

Another aspect of the invention relates to enzymatic propagation of a nucleic acid sequence on at least two surfaces, comprising:

(a) immobilizing a first primer on at least one first surface of a substantially solid phase;

(b) administering a solution of nucleic acids comprising complementary fragments to the first primer;

(c) binding of complementary fragments to the first primer;

(d) extending the first primer at its 3' end, corresponding to the complementary fragment by means of a polymerase;

(e) releasing the complementary fragments;

(f) attaching a second primer to the 3' end of the extended nucleic acid;

(g) extending the second primer at its 3' end by means of a polymerase;

(h) transferring the second primer to another surface and immobilization of the extended primer; and

(i) attaching another first primer to the 3' end of the second extended primer for further extending of the first primer, wherein the steps of (b) to (h) are repeated numerous times for exponential amplification of nucleic acid sequence.

The second primer as set forth in step (h) is transferred to a second surface which is spatially separated from the first, and the transfer is achieved by the application of an electrical field. Preferably, the electrical field is applied between the first and second surface.

Further amplification stages may occur following the transferal of the second primer to the second surface comprising the following:

(j) extending this first primer to its 3' end, corresponding to the complementary fragment, by means of a polymerase;

(k) transferring of the extended primer to the first or another surface and immobilization of the extended primer thereon; and

(l) attaching of another second primer to the 3' end of the extended first primer.

A further aspect of the invention relates to a method for copying nucleic acids from a first to a second surface, comprising:

(a) immobilizing of nucleic acids through a reaction on a carrier surface;

(b) producing a double-stranded molecule by a method selected from the group consisting of hybridization of complementary single strands, chemical or enzymatic ligation of complementary fragments and chemical or enzymatic extension of complementary primers; and

(c) transferring of complementary strands to a second surface with immobilization of the complementary strands thereon.

The transfer of the complementary strands to the second surface may be accomplished by the application of an electrical field, in which the electrical field is applied between the first and second surface.

Preferably, the nucleic acids immobilized on the solid carrier are arranged two-dimensionally and are transferred in this order, while retaining site information. The solid phase material is selected from organic or inorganic material or from a hybrid of these materials, and preferably represents a two- or three-dimensional matrix. Immobilization of the nucleic acids and complementary strands occurs through covalent or non-covalent binding.

In still another aspect, the nucleic acids, ligands, receptors or their derivatives are provided with a detectable label. Generally, any molecular moiety capable of detection may be utilized including, by way of example, without limitation radioisotopes, stable isotopes, enzymes, immunoreactive

compounds, fluorescence or luminescence chemicals, chromophores, metals or charged particles.

The present invention may be applicable for several different amplification methods including, cloning genomic fragments of DNA, cDNA and RNA, subcloning following restriction-digesting, strengthening an immunological ligand/receptor pair, strengthening the ligand signal, sorting adjacent fragments by using hybridization techniques (chromosome walking), and copying of gene chips.

The term "amplification" is used in the broad sense to mean creating a product which may include, by way of example, additional target molecules, or target-like molecules or molecules complementary to the target molecules, which molecules are created by virtue of the presence of the target molecule in the sample. In a situation where the target is a nucleic acid, an amplification product can be made enzymatically with an agent for polymerization, such as with DNA or RNA polymerases or transcriptases.

The above-discussed embodiments of the present invention will be described further hereinbelow. When the word "invention" is used in this specification, the word "invention" includes "inventions", that is the plural of "invention". By stating "invention", the Applicant does not in any way admit that the present application does not include more than one patentably and non-obviously distinct invention, and maintains that this application may include more than one patentably and non-obviously distinct invention. The Applicant hereby asserts that the disclosure of this application may include more than one invention, and, in the event that there is more than one invention, that these inventions may be patentable and non-obvious one with respect to the other.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1A describes a general process for the propagation of ligands and receptors on two surfaces.

FIG. 1B shows the process in accordance with FIG. 1, having an intermediate layer placed between the two surfaces.

FIG. 2 describes a general process for propagating ligands and receptors on a surface.

FIG. 3A describes a process for the enzymatic propagation of nucleic acids on two surfaces.

FIG. 3B shows the process in accordance with FIG. 3, having an intermediate layer placed between the two surfaces.

FIG. 4 shows a basic process for the enzymatic propagation of nucleic acids on only one surface.

FIG. 5 shows a process for copying nucleic acids onto a second surface.

FIG. 6 is a diagram showing two systems coordinates containing numerous fields which are intended to illustrate two surfaces.

FIG. 7 shows a process for synthesizing genes and genomes, and for recombination.

FIG. 8 shows a process for the selective mutagenesis of nucleic acids site-directed mutagenesis.

FIG. 9 shows a process for cloning and sequencing genomic fragments.

FIG. 10 shows a process for the functional analysis of genomic fragments.

FIG. 11 shows a process for the parallel quantifying of the gene expression.

FIG. 12 shows the use of a process as a means of improving the signal-to-noise ratio in the detection process.

FIG. 13 shows a process used for the functional location of proteins.

FIG. 14 shows a process used to screen combinatory protein libraries.

DETAILED DESCRIPTION AND PREFERRED EMBODIMENTS

The present invention is a method for cloning and copying ligand/receptor matrices by exponentially amplifying the components of the matrices and immobilizing the components of the matrices on at least one surface and separating the components by means of an electric field while retaining site information.

Nucleotide as used herein means a monomeric unit of DNA or RNA consisting of a sugar moiety (pentose), a phosphate, and a nitrogenous heterocyclic base. The base is linked to the sugar moiety via the glycosidic carbon (1' carbon of the pentose) and that combination of base and sugar is a nucleoside. When the nucleoside contains a phosphate group bonded to the 3' or 5' position of the pentose it is referred to as a nucleotide. The nucleotides are adenine, thymine, cytosine, guanine and uracil.

Base Pair (bp) as used herein means a partnership of adenine (A) with thymine (T), or of cytosine (C) with guanine (G) in a double stranded DNA molecule. In RNA, uracil (U) is substituted for thymine.

Nucleic acid as used herein means a polymer of nucleotides, either single or double stranded.

Oligonucleotide as used herein makes reference to primers, probes and nucleic acid fragments or segments to be synthesized by ligation of oligonucleotides is defined as a molecule comprised of two or more deoxyribonucleotides or ribonucleotides, preferably at least 5. Its exact size will depend on many factors, which in turn depends on the ultimate conditions of use.

Gene as used herein means a nucleic acid whose nucleotide sequence codes for an RNA or polypeptide. A gene can be either RNA or DNA.

Complementary Nucleotide Sequence as used herein means a sequence of nucleotides in a single-stranded (ss) region of DNA or RNA capable of hybridizing to another single-stranded region for a length of time sufficient to permit the desired reaction, e.g., a ligation reaction or a primer extension reaction.

Conserved as used herein means a nucleotide sequence is conserved with respect to a preselected (reference) sequence if it non-randomly hybridizes to an exact complement of the preselected sequence.

Hybridization as used herein means the pairing of substantially complementary nucleotide sequences (strands of nucleic acid) to form a duplex or heteroduplex by the establishment of hydrogen bonds between complementary base pairs. It is a specific, i.e. non-random, interaction between two complementary polynucleotides that can be competitively inhibited.

The present invention relates to a method of amplification comprising:

(a) immobilizing a first primer on a first surfaces of a substantially solid phase;

(b) administering a solution of nucleic acids comprising complementary fragments to the first primer;

(c) binding of complementary fragments to the first primer;

(d) extending the first primer at its 3' end, corresponding to the complementary fragment by means of a polymerase;

(e) releasing of the complementary fragments;

(f) attaching a second primer to the 3' end of the extended nucleic acid;

(g) extending the second primer at its 3' end by means of a polymerase;

(h) transferring the second primer to another surface and immobilization of the extended primer thereon; and

(i) attaching another first primer to the 3' end of the second, extended primer.

Specifically, the term "primer" as used herein refers to a molecule comprised of two or more deoxyribonucleotides or ribonucleotides, preferably more than three, which is capable of acting as a point of initiation of synthesis when placed under conditions in which synthesis of a primer extension product which is substantially complementary to a nucleic acid strand is induced, i.e., in the presence of nucleoside triphosphates and an agent for polymerization such as DNA polymerase and at a suitable temperature and pH. The primer is preferably single stranded for maximum efficiency in amplification, but may alternatively be double stranded. If double stranded, the primer is first treated to separate its strands before being used to prepare extension products. Preferably, the primer is an oligodeoxyribonucleotide. The primer may be sufficiently long to prime the synthesis of extension products in the presence of the inducing agent for polymerization. The exact lengths of the primers will depend on many factors, including temperature, buffer, nucleotide composition and source of primer. For purposes herein, an oligonucleotide primer typically contains from about 5 to about 50 nucleotides, and preferably from about 5 to about 15.

The primers herein are selected to be "substantially" complementary to each strand of the specific sequence to be amplified. This means that the primers must be sufficiently complementary to hybridize with their respective strands under conditions which allow the agent for polymerization to perform, i.e, the primers have sufficient complementarity with the sequence of the strand to be amplified to hybridize therewith and thereby form a template for synthesis of the extension product of the other primer. Preferably, the primers have exact complementarity with the strand.

The oligonucleotide primers may be prepared using any suitable method. In an automated embodiment, diethylphosphamidites are used as starting materials and may be synthesized as described by Beaucage et al. Tetrahedron Letters (1981), 22:1859–1862. One method for synthesizing oligonucleotides on a modified solid support is described in U.S. Pat. No. 4,458,066. It is also possible to use a primer which as been isolated from a biological source (such as a restriction endonuclease digest).

The role of the solid phase surface (also referred to in the following as a "carrier") in the present invention includes maintaining a separation between the complementary matrices which would form stable duplexes if in solution. Suitable carrier materials consist of organic or inorganic materials or a hybrid of these materials. Organic carrier materials consist of polymers on a sugar basis, preferably agarose, cellulose, and suitable derivatives or technical polymers such as polystyrene, polyacrylate, polyacrylnitril, polyalkene or graft copolymers (e.g. PS PEG, PAN-PEG, PAN-PAG etc.), as well as electrically conductive polymers (e.g. polyvinylpyrrol). Examples of suitable inorganic carrier materials include glass or metals, with particular importance being attached to gold surfaces (as a result of the gold thiolate interaction) and semiconductor surfaces.

A preferred carrier support comprises a sheet that has surfaces with alignment features to allow the precise positioning of the nucleic acids to facilitate the transference of polymerized molecules to another support having the surface defined in the same grid type pattern thereby retaining site location. During the respective transfer stage within the respective method sequence, the scale can be reduced and/or increased, while retaining the site information. The supports may be filters, fibers, membranes, sheets and the like. It is preferable, but not essential, that the surfaces should be arranged so that they are coplanar to one another.

Bonds formed between the primers and support material may be either covalent or non-covalent, with non-covalent bonds encompassing both ionic and non-ionic binding systems, and in particular members of immunological pairs of bonds such as avidin/streptavidin and antigen antibodies. The primers may be immobilized onto the solid support or carrier in any art-recognized way. A commonly used means is to provide a biotin label on the primer for binding to a streptavidin-coated support. Suitable binding linkers should not form undesirable interactions with other factors occurring in the system. In particular, in the hybridization of a primer, no interaction with surface zones containing the template should occur. This imposes the requirement for a controllable binding chemistry that can be influenced by external conditions.

Faulty immobilization can be prevented by using pairs of primers (see below) instead of an "activatable reactive primer". It should be possible to immobilize the extension products of these pairs of primers orthogonally. In this case orthogonally means that no binding points are available for a primer that is hybridized on a template, but that binding points are available following translocation of the primer extension product to the opposite surface. The way in which the primer is added, and the way in which the reaction is allowed to take place must take this aspect into account. As defined by this invention, activatable reactive primers are considered to be those primers that function reactively and whose reactivity can be influenced by the choice of suitable external conditions.

Any source of nucleic acid, in purified or non-purified form, can be utilized as the starting nucleic acid or acids, provided it contains or is suspected of containing the specific nucleic acid sequence associated with that to be detected and amplified. One may select the sequence being amplified from among the regions that are substantially conserved among the biological material of interest. The method of the present invention may employ, for example, DNA or RNA, including messenger RNA, and the DNA or RNA may be single stranded or double stranded. In the event that RNA is to be used as a template, enzymes and/or conditions optimal for reverse transcribing the template to DNA should be utilized. In addition, a DNA-RNA hybrid which contains one strand of each may be utilized. A mixture of any of these nucleic acids may also be employed, or the nucleic acids produced from a previous amplification reaction herein using the same or different primers may be so utilized.

The target nucleic acid sequence to be amplified or copied may be only a fraction of a larger molecule or can be present initially as a discrete molecule, so that the specific sequence constitutes the entire nucleic acid. It is not necessary that the sequence to be amplified be present initially in a pure form; it may be a minor fraction of a complex mixture, such as a portion of the virus-encoding gene contained in a whole human DNA. The starting target nucleic acid may contain more than one desired specific nucleic acid sequence which may be the same or different. Therefore, the present method is useful not only for producing large amounts of one specific nucleic acid sequence, but also for amplifying

simultaneously more than one different specific nucleic acid sequence located on the same or different nucleic acid molecules.

In the process described herein, the copies of nucleic acids thus produced may be identical or complementary to the initial sequence (matrix). "Complementary" in the case of this invention is taken to mean that the copy of the matrix differs from the initial matrix, whereas the copy of this copy is identical with the initial matrix. If necessary, these will be referred to in abbreviated form in the following as the "(+) strand" and "(−) strand". The reactions take place in the same reaction vessel. Nucleic acids are taken to mean both D- and L-nucleic acids (mirror mers), and any kind of modifications to them.

The nucleic acid(s) may be obtained from any source, for example, natural DNA or RNA from higher organisms such as animals. DNA or RNA may be extracted from a bodily sample such as blood, tissue material such as chorionic villi, or amniotic cells by a variety of techniques such as that described by Maniatis et al., Molecular Cloning (1982), 280–281. Preferably, the first nucleic acid has a size of approximately 1 to 10 Kb. Larger nucleic acid can be readily digested by enzymes or mechanically fragmented.

Any specific nucleic acid sequence can be copied and amplified by the methods of present invention. It is only necessary that a sufficient number of bases at both ends of the sequence be known in sufficient detail so that at least one oligonucleotide primer can be prepared that will hybridize to different strands of the desired sequence and at relative positions along the sequence such that an extension product synthesized from one primer, when it is separated from its template (complement), can serve as a template for extension of the other primer into a nucleic acid of defined length. The greater the knowledge about the bases at both ends of the sequence, the greater can be the specificity of the primers for the target nucleic acid sequence, and thus the greater the efficiency of the process. It will be understood that the word primer as used hereinafter may refer to more than one primer, particularly in the case where there is some ambiguity in the information regarding the terminal sequence(s) of the fragment to be amplified. For instance, in the case where a nucleic acid sequence is inferred from protein sequence information a collection of primers containing sequences representing all possible codon variations based on degeneracy of the genetic code will be used for each strand. One primer from this collection will be substantially conserved with the end of the desired sequence to be amplified.

A nucleic acid sequence is produced by using the target nucleic acid containing that sequence as a template. If the target nucleic acid sequence of the sample contains two strands, it is necessary to separate the strands of the nucleic acid before it can be used as the template. This strand separation can be accomplished using any suitable denaturing conditions, including physical, chemical or enzymatic means, the word "denaturing" used herein to include all such means. One physical method of separating the strands of the nucleic acid involves heating the nucleic acid until it is denatured. Typical heat denaturation may involve temperatures ranging from about 80° C. to about 105° C., for times ranging from about 1 to 10 minutes. Strand separation may also be induced by an enzyme from the class of enzymes known as helicases or the enzyme RecA, which has helicase activity and in the presence of riboATP is known to denature DNA. The reaction conditions suitable for separating the strands of nucleic acids with helicases are described by Kuhn Hoffmann-Berling, CSH-Quantitative Biology, 43:63 (1978), and techniques for using RecA are reviewed by C. Radding, Ann. Rev. Genetics, 16:405–37 (1982).

If an appropriate primer is immobilized on a support surface and a single-stranded target nucleic acid sequence (acting as a template) is added to a system then a primer extension product is synthesized in the presence of a polymerization agent, and the four nucleoside triphosphates described below. The product will be at least partially complementary to the single-stranded target nucleic acid and will hybridize with the target nucleic acid strand to form a duplex of substantially equal length strands that may then be separated into single strands as described above to produce two single separated complementary strands.

The techniques used for amplifying and thereafter detecting the product are described in detail U.S. Pat. Nos. 4,683,195 and 4,683,202, the entire disclosures of which are incorporated herein by reference. In general, the amplification process involves an enzymatic chain reaction for preparing, in exponential quantities relative to the number of reaction steps involved, a specific nucleic acid sequence, given that the ends of the required sequence are known in sufficient detail that oligonucleotide primers can be synthesized which will hybridize to them, and that a small amount of the sequence is available to initiate the chain reaction. Preferably, one primer is complementary to the negative (−) strand and the other is complementary to the positive (+) strand.

The synthesis is performed under conditions allowing hybridization of primers to templates to occur. Generally, hybridization occurs in a buffered aqueous solution, preferably, at a pH of about 7 to about 9, most preferably about 8. Preferably, a molar excess of the primers is added to a buffer solution for immobilization on the support before the template strands are introduced. It is understood, however, that the amount of template strand may not be known if the process herein is used for diagnostic applications, so that the amount of primer relative to the amount of template (complementary) strand cannot be determined with certainty. As a practical matter, however, the amount of primer added will generally be in molar excess over the amount of complementary strand (template) when the sequence to be amplified is contained in a mixture of complicated long-chain nucleic acid strands. A large molar excess is preferred to improve the efficiency of the process.

Deoxyribonucleoside triphosphates dATP, dCTP, dGTP and dTTP are also added to the synthesis mixture in adequate amounts to provide sufficient building blocks for synthesizing an extended primer sequence and the resulting solution is heated to about 90° C. to about 110° C. for about 1 to 15 minutes, preferably from 1 to 4 minutes. After this heating period the solution is allowed to cool to room temperature, which is preferable for the primer hybridization. To the cooled mixture is added an appropriate agent for effecting the primer extension reaction (called herein "agent for polymerization"), and the reaction is allowed to occur under conditions known in the art. The agent for polymerization may also be added together with the other reagents if it is heat stable. This synthesis reaction may occur at from room temperature up to a temperature above which the agent for polymerization no longer functions. Thus, for example, if DNA polymerase is used as the agent, the temperature is generally no greater than about 40° C. Most conveniently the reaction occurs at room temperature.

The agent for polymerization may be any compound or system which will function to accomplish the synthesis of primer extension products, including enzymes. Suitable enzymes for this purpose include, for example, *E. coli* DNA

polymerase I, Klenow fragment of *E. coli* DNA polymerase I, T4 DNA polymerase, other available DNA polymerases, polymerase muteins, reverse transcriptase, and other enzymes, including heat-stable enzymes (i.e., those enzymes which perform primer extension after being subjected to temperatures sufficiently elevated to cause denaturation), that will facilitate combination of the nucleotides in the proper manner to form the primer extension products which are complementary to each nucleic acid strand used as a template. Generally, the synthesis will be initiated at the 3' end of each primer and proceed in the 5' direction along the template strand, until synthesis terminates. There may be agents for polymerization, however, which initiate synthesis at the 5' end and proceed in the other direction, using the same process as described above.

The newly synthesized strand and its complementary nucleic acid strand (template strand) will form a double-stranded molecule under the hybridizing conditions described above and this hybrid is used in the succeeding steps of the process. In the next step, the double-stranded molecule is subjected to denaturing conditions using any of the procedures described above to provide single-stranded molecules if the target sequence is present. Denaturation of the complementary strands can be performed by subjecting the system to heat, for example 90° C. to 110° C. for about 2 to 15 minutes, or highly alkaline conditions, such as by the addition of sodium hydroxide. Denaturation can also be accomplished by adding organic acids, nucleic acid binding proteins or enzymes which promote denaturation.

Preferably, the release of the first nucleic acid (template strand) allows further amplification of the desired product by allowing the extended nucleic acid sequence to act as a template for further hybridization reactions to form a second hybridization product. New nucleic acid may be synthesized on the single-stranded first hybridization product which is the extended nucleic acid sequence. Additional agent for polymerization, nucleotides and primers are added, if necessary, for the reaction to proceed under the conditions prescribed above. Preferably a second primer is attached to the 3' end of the extended nucleic acid sequence and the second hybridization product will mimic the sequence of the first nucleic acid template strand. The 5' end of the second primer, attaches to the first extended nucleic acid, so that the 3' end of the primer can be extended in the next hybridization process. The second primer will include a functional 5' end group to facilitate immobilization on a second solid support in subsequent reactions.

Upon completion of the second hybridization product, the double stranded nucleic acid molecules comprising the first and second extended nucleic acid sequences is separated by any denaturing process as discussed hereinabove. Upon separation of the double stranded nucleic acid molecules, the first solid phase support is introduced to a standard electrophoresis chamber comprising a second solid phase surface. Application of an electric field, in the electrophoresis chamber causes the newly hybridized nucleic acid strands (after separation) to migrate towards the second solid phase surface for an immobilization reaction.

The present inventive method for amplifying target sequences comprises subjecting the synthesized nucleic acid strands in a suitable medium to an electric field generated by at least two electrodes arranged which serve as driving electrodes. A standard horizontal gel apparatus is described in T. Maniatis, E. F. Fritsch, and J. Sambrook, Molecular Cloning: A Laboratory Manual, 153 (Cold Spring Harbor Laboratory, 1982) which comprises a positive and negative electrode, submerged in solution at opposite ends of a buffer tank.

The driving electrodes have a potential difference imposed across them and this determines the general direction of the electric field. The electric field effectively controls the migration of the non-immobilized strand of nucleic acid effecting separation of nucleic acids strands. The field may be uniform or non-uniform and alternating/non-alternating fields may also be employed for modifying the resolution, migration path and rate of separation.

The electrophoresis is carried out in a gel conventionally prepared using agarose or polyacrylamide as the gelling agent, although other gelling agents may find use. In carrying out the method of the subject invention an electric field of a desired shape is created in the gel. The field is maintained during the electrophoresis and may be uniform or non-uniform, alternating or non-alternating. The gel is supported in a reservoir containing a buffered medium and located between the electrodes. The solid phase support may be loaded near one end of the gel. The gel is positioned within the electrode array to provide the proper migration direction for the charged sequence.

In practice, the electrodes are suspended in a buffer tank and immersed in a buffer solution. The buffer solution can be circulated and cooled in conventional fashion. The first solid phase support is positioned in the chamber so that the direction of travel of the free nucleic acid is towards the second solid phase support. The potential across the electrodes is determined by the size of the free nucleic acid sequence that are moved through the electric field. The velocity of the free nucleic acids or biological materials is dependent on the strength electric field, the net charge on the free nucleic acid and friction resistance. The isoelectrical point should be determined for the free nucleic acid to determine the appropriate pH in the solution for increased velocity because the net charge is dependent upon the pH of the solution. Further, the primers may be provided with charged groups to increase the velocity in the electric field. The electric field may be continuous or pulsed and pulse times and voltages will vary with the size of the molecules to be resolved, pulse times being longer with larger biological molecules. Pulse times will generally be in the range of 1 second to several hours, more usually in the range of 5 seconds to 60 minutes. The voltages will generally be in the range of about 50 to about 500 volts which will effectively cause the migration of the free nucleic acid sequence in the direction of the second solid phase surface.

Upon contact of the second primer with the second solid phase surface, it is immobilized thereon by attaching of the 5' end of the second primer with a functional grouping on the second solid phase surface.

The steps of extension product synthesis, denaturing and application of an electric field can be repeated as often as needed to amplify the first target nucleic acid sequence to the extent necessary for detection. The amount of the synthesized nucleic acid sequence produced will accumulate in an exponential fashion.

When it is desired to produce more than one synthesized extended nucleic acid sequence from a mixture of target nucleic acids, the appropriate number of different oligonucleotide primers are utilized. For example, if two different nucleic acid sequences are to be synthesized, four primers are utilized. Two of the primers are specific for one of the specific nucleic acid sequences and the other two primers are specific for the second nucleic acid sequence. In this manner, each of the two different sequences can be produced exponentially by the present process.

The invention also applies in particular to a process for the enzymatic propagation of nucleic acids on at least two

surfaces, and incorporating one or more of the following amplification cycles

(a) immobilizing a first primer on a first surface of a solid phase;

(b) adding a solution of nucleic acids and binding of complementary fragments to the first primer;

(c) extending the first primer at its 3' end, corresponding to the complementary fragment, by means of a polymerase;

(d) releasing the complementary fragments by denaturing;

(e) adding a second primer to the 3' end of the extended nucleic acid;

(f) extending the second primer at its 3' end by means of a polymerase;

(g) releasing the second primer extended nucleic acid and transferring to an additional surface for immobilization thereon;

(h) adding additional first primer for extending the additional first primer to the 3' end of the second extended primer.

The present method is useful for the propagation of biological material on so-called "gene chips". Basically it is immaterial (as it is in any case in the method involved in this invention) whether the 5' end, the 3' end or an internal position within the sequence is used as the starting point for the link with the surface.

The knowledge already exists regarding the attachment or synthesis of a large number of immobilized polymer compounds onto an object carrier as a way of demonstrating the action of selectively binding compounds on such carriers (Fodor et al., Science 251, 767–773, 1991; U.S. Pat. Nos. 5,510,270, 5,489,678, 5,445,934, and 5,424,186). However, the process of producing such arrays of probes must be preceded by the creation and use of lithographic masks, and the initial monomer compounds must be provided with light-sensitive protective groups. The synthesis cycle in the case of peptide synthesis requires at least 20 such masks for each cycle, i.e. for n cycles the number of masks required is n×20; in the synthesis of oligonucleotides 4 such lithographic masks are required, i.e. if there are n cycles, n×4 masks are required. These lithographic masks are needed to permit illumination at defined spatial points on the array, while preventing illumination at other points on the array. A light-sensitive protective group is split off at the defined, illuminated points, and this enables a reactive group to be released, to which a new monomer building block of the polymer can subsequently bind. Such arrays are built up through the repeated application of individual masks and the multiple repetition of coupling processes. Until now complicated and extremely expensive lithographic methods have been needed to produce such gene chips (see also U.S. Pat. No. 5,700,637).

The methods of the present invention unexpectedly provide a simple alternative to such methods described in the prior art, and moreover offer much greater levels of efficiency and precision. Because of the wealth of information that can be held on such gene chips, it is even possible to prepare entire genetic databases or libraries for screening purposes. The purpose of this invention is to arrange the nucleic acids two-dimensionally on the fixed carrier, and to transfer them in accordance with this arrangement. As defined by this invention the two-dimensional arrangement can also be regarded as a "disarray". Especially in the case of large libraries there inevitably no specific allocation with regard to the way in which individual molecules bind to particular sites. In a spatial sense, however, when a transfer takes place the non-allocated molecules of the library are transferred along with their original, site-specific information. We then have a "disarray" which, at the same time, is transferred in an organized form, while retaining site-specific information.

Additional applications for the present invention include the production of gene chips for diagnostic purposes in both human and veterinary medicine.

Immune reactions in the form of immunoassays or RIAs (antibody-antigen reactions) can also be carried out using the methods referred to in this invention.

The invention offers previously unsuspected advantages for all known, relevant methods, i.e. conventional, medico-diagnostic and biochemical/biotechnological/genetic engineering, and consequently it opens up many fields of application. A few of these applications are set forth below.

The method referred to in this invention can be used for the qualitative and quantitative detection of DNA and RNA molecules. This method also enables complex genetic polymorphism and multiple alleles to be analyzed simultaneously. The propagation of the DNA or RNA molecules on solid phases enables the avoidance of primer artifacts (e.g. primer-dimers). The fact that the sample to be quantified is only introduced at the start of the process and that subsequently all propagation products are firmly attached to the surfaces by specific bonds can be seen as yet another advantage. In this way the frequent interference from signals produced by impurities during PCR diagnosis is avoided. Surface purity can be improved by the electrostatic rejection of non-specifically bonded DNA or RNA molecules. The method can therefore be applied in functional genomics and pharmacogenomics (see Oliver et al. Trans Biotechnol. 16, 373–378 (1998); Housman and Ledley, Nature Biotechnology 16, 492–493 (1998)).

The methods referred to in this invention can be used to detect differential gene expressions. Furthermore the process can be combined with state of the art methods such as differential display RT-PCR (DDRT-PCR), serial analysis of gene expression (SAGE) or differential hybridization (Wan et al., Nature Biotechnology 14, 1685–1691 (1996)).

For rapid qualitative and quantitative detection the methods of the present invention can be combined with known, state-of-the-art sensoring methods, e.g. surface plasmon resonance sensors, evanescent field sensors, faser optic sensors, grating couplers or RIFS (reflector-interferometer spectroscopy) (Scheller et al., Frontiers in Biosensorics, Birkhäuser Verlag Basel (1997)).

The compilation of gene and genome libraries in accordance with this invention can be combined with the known, state of the art process for the ligation of adapters or linkers. One particular advantage of this process is that, when two different primers are used for surface immobilization, only molecules with two different adapters or linkers are propagated. In this way the respective, complementary, individual strands on the surfaces are propagated separately.

Moreover, a combination is also possible with processes which permit site-specific immobilization and the re-sorting of surface-bonded molecules. These include firstly the methods based on arrays of electrodes in which one or more specific microelectrodes can be selectively triggered. In such cases, the arrays of electrodes may be made up using semiconductor chips, for example, such as those developed by the Nanogen Company (www.nanogen.com). Other methods involve the use of scanning techniques, in which piezoelectric elements are used to ensure extremely precise lateral addressing down to the sub-nanometer range. The preferred approach involves the use of scanning electro-

chemical microscopy (SECM) for electrochemical deposition of molecules. This can also be used in association with electrochemical probes for detection purposes. Methods such as atomic force microscopy (AFM) are also suitable for the lateral translocation of individual molecules.

The gene and genome libraries compiled in accordance with the methods of the present invention can be sequenced using the known state-of-the-art method, e.g. sequencing by chemical splitting, sequencing by hybridization, sequencing by capillary electrophoresis or MALDI mass spectrometry (see Adams, Fields, Venter in: Automated DNA Sequencing and Analysis, Academic Press, 1994). The gene and genome libraries thus compiled can be used to allocate DNA and RNA-binding factors. For example the specific binding sites for transcription activators or repressors can be detected simultaneously. The gene and genome libraries used may be either single or double stranded.

The methods of the present invention may be used for gene and genome synthesis and for the recombination of genetic material. The method described in FIG. 7 (see below) permits the connection of any number of fragments possessing only partial complementarity. In particular it enables open scanning grids to be provided with suitable starting signals for biological expression systems. In addition the process can be coupled with in-vitro transcription and in-vitro translation, because transcription and translation products can be created while retaining site information. Also, it is possible to test the spatial arrangement of these products in the way that they interact with other factors. This enables functional allocations to be detected simultaneously which, in the current state of the art, is only found in complicated systems such as the "Two Hybrid-System" (Fredericson, Curr. Opin. Biotechnol. 9, 90–96 (1998)). Thus, using the present methods, new effective pharmacological sites can be found, or new diagnostic strategies developed.

Moreover, the process can be combined with known, state of the art methods for finding functional molecules from combinatorial libraries. For example, a target molecule can be immobilized on a first surface. The non-binding RNA or DNA molecules can be separated following contact with a combinatorial nucleic acid library. The binding molecules obtained from the combinatorial library are transferred to a second surface, using the process to which this invention refers. They are provided with information about the site and then propagated. Details about the composition of the binding molecules can be obtained through sequencing.

In addition a cyclical procedure can be used to achieve evolutive optimization of the binding molecules. The molecules that have already been selected are again brought into contact with the immobilized target molecules and are again propagated. Faulty propagation leads to the creation of secondary molecules, in some cases with improved binding properties. The population of the sequence derived from the primary sequence through mutation is referred to by Eigen as a quasi-sequence. The process to which this invention refers is distinguished by the fact that the members of the quasi-species are spatially co-localized, giving a lateral dimension to the concept of the quasi-species.

By applying increasingly stringent conditions, e.g. reduction of the target concentration and increased flushing, it is possible to attain systematic optimization of the functional properties. The selection pressure can also be achieved through increasing electrostatic repulsion. Moreover the process can also be combined with known, state of the art methods for the selection of catalytic nucleic acids (Tarosow et al., Nature 389, 54–57 (1997)).

Consequently, the methods of the present invention can be used in particular for cloning genomic fragments of DNA, cDNA and RNA, especially using replicative polymerases, e.g. polymerase III derived from *Escherichia coli*. Advantageously, errors are avoided in the replication process. Moreover, subcloning is available following restriction-digesting. Subcloning has a part to play in the sequencing of large genomic fragments and the present methods achieve considerable time savings during sorting the large libraries of nucleic acids, for example by the use of subcloning techniques such as "shotgun cloning" and the creation of selective deletion variants, e.g. exonuclease III treatment (see Adams, Fields, Venter, in: Automated DNA Sequencing and Analysis, Academic Press, 1994).

Further, the present invention has advantages for sorting adjacent fragments through the use of hybridization techniques ("chromosome walking").

Using the instant methods in order to copy gene chips offers previously unsuspected possibilities. For example, gene chips can be used with a library of viral antigens as the source material. The reaction with a patient's blood and the action of binding the antibodies that the blood contains enable an immune reaction to be detected. The process can be repeated any number of times, which in turn strengthens the signal. To an extent a linear propagation thus takes place, weak immune signals can be identified. This is useful, for example, in the diagnosis of AIDS, which, as is commonly known, is difficult to detect at the start of the infection. It is also possible to identify diseases or allergies in their early stages. The present methods are also suitable for the exponential propagation which may occur when using an immobilized antigen. In addition, using semi-conductor technology and/or micromanipulation ("align techniques") a sorting process can be carried out on gene chips. It is also possible to provide the charged nucleic acids with a positively charged group of headings, which predominates in the overall charge. This enables the molecules to be aligned in the electrical field, permitting a high charge density on the chips (DNA/RNA).

FIG. 1A describes a general process for the propagation of ligands and receptors on two surfaces; the individual stages are:

(1) as a result of a binding reaction, a ligand is immobilized on the surface of a solid carrier;

(2) the ligand binds a receptor;

(3) the receptor is transferred to a second surface by the application of an electrical field whereon the receptor is immobilized by a reaction;

(4) a free ligand is added to the immobilized receptor; and

(5) the ligand is transferred to the second surface (with the aid of an electrical field) for immobilization thereon.

The process may be carried out or repeated any number of times.

FIG. 1B shows the process in accordance with FIG. 1A, but here an intermediate layer is placed between the two substantially solid phase surfaces. This intermediate layer is chosen from among a group comprising a gel, a membrane, a polymer, a ceramic and/or a so-called capillary tube array. When an electrical field is applied, the non-immobilized molecule travels via the intermediate layer to the second surface, where it is immobilized. The intermediate layer can be permeated by nucleic acids and/or ligands/receptors.

FIG. 2 describes a general process for propagating ligands and receptors on a surface; the individual stages are:

(1) a ligand is immobilized on the surface of a solid carrier by a binding reaction;

(2) the ligand binds a receptor;

(3) the receptor is transferred to the surface by the application of an electrical field to be immobilized thereon by a reaction;

(4) a free ligand is added to the immobilized receptor; and

(5) the ligand is immobilized on a surface by means of a reaction.

The process may be repeated any number of times.

FIG. 3A describes a process for the enzymatic propagation of nucleic acids on two surfaces, with the individual stages of the process being shown in fast motion as follows:

(1) a first primer A is immobilized by a binding reaction on a solid carrier surface;

(2) the primer A bonds complementary fragments from a solution of nucleic acids;

(3) the primer A is extended by a polymerase at its 3' end;

(4) the complementary fragments are released;

(5) a second primer B is added to the 3' end of the extended nucleic acid;

(6) the primer B is extended by a polymerase at its 3' end;

(7) the extended, non-immobilized primer B is transferred to a second surface by applying an electrical field, wherein the two surfaces having opposed polarities, and the extended primer B is immobilized thereon by a binding reaction;

(8) an additional primer A is added to the 3' end of the extended primer B;

(9) the primer A is extended by a polymerase at its 3' end;

(10) the extended primer A is transferred to an additional surface by applying an electrical field, with the two surfaces having opposed polarities, and the extended primer A is immobilized there by a reaction; (in this case, unlike in stage (7) the polarities are reversed) (11) another primer B is added to the 3' end of the extended primer A.

The process is then repeated any number of times, preferably by means of a cycle of polarity reversals.

FIG. 3B shows the process in accordance with FIG. 3A, but here an intermediate layer is placed between the two surfaces. This intermediate layer is chosen from among a group comprising a gel, a membrane, a polymer, a ceramic and/or a so-called capillary tube array. When an electrical field is applied, the non-immobilized molecule travels via the intermediate layer to the second surface, where it is immobilized.

FIG. 4 shows a basic process for the enzymatic propagation of nucleic acids on only one surface; the individual stages are as follows:

(1) a first primer A is immobilized by a reaction on the surface of a solid carrier;

(2) the primer A binds complementary fragments from a solution of nucleic acids;

(3) the primer A is extended by a polymerase at its 3' end;

(4) the complementary fragments are released;

(5) a second primer B is added to the 3' end of the extended nucleic acid;

(6) the primer B is extended by a polymerase at its 3' end;

(7) the extended, non-immobilized primer B is transferred to the surface of the solid carrier by applying an electric field and bonded thereon by means of an irreversible reaction, the charged primer B being conducted along the field; and

(8) an additional primer A is added to the 3' end of the extended primer B.

The process can be continued any number of times.

In contrast, FIG. 5 shows a process for copying nucleic acids onto a second surface; the individual stages are as follows:

(1) immobilization of nucleic acids by a reaction with the surface of a solid carrier;

(2) production of double-stranded molecules through the hybridization of complementary single strands; alternatively chemical or enzymatic ligation of complementary fragments may take place, or chemical or enzymatic extension of complementary primers; and

(3) transfer of complementary strands to a second surface, where they are immobilized; this is done preferably by applying an electrical field, the two surfaces being of opposing polarities.

FIG. 6 is a diagram showing two systems of coordinates containing numerous fields which are intended to illustrate two surfaces, such as those used, for example, in the case of gene chips or membranes. The lower surface contains the information to be copied, which is transferred to the upper surface by the application of an electrical field (not shown) corresponding to the process described in FIG. 3A. It is then immobilized on this surface.

FIG. 7 shows a process for synthesizing genes and genomes, and for recombination; the individual stages are as follows:

(1) a primer A is immobilized by a reaction on the surface o of a solid carrier;

(2) the primer A binds complementary fragments from a solution of nucleic acids;

(3) the primer A is extended by a polymerase at its 3' end;

(4) the complementary fragments are released;

(5) a second fragment, complementary with the 3' end of the extended primer A, is added, and here partially complementary fragments with overhanging 3' ends are sufficient;

(6) the extended primer A is elongated;

(7) the complementary fragments are released.

Stages (5)–(7) can be repeated any number of times;

(8) a second primer B is added at the 3' end of the extended primer A;

(9) the primer B is extended by a polymerase at its 3' end; and

(10) as described in FIG. 1A, Stage 7, the extended primer B is transferred to a second surface. The advantage of this step is that incompletely extended primer A molecules are separated out. The extended primer B can again be used for extension reactions.

FIG. 8 shows a process for the selective mutagenesis of nucleic acids (site-directed mutagenesis); the individual stages are as follows:

(1) a nucleic acid A immobilized by a reaction on the surface of a solid carrier; the nucleic acid may be one of the products of the processes referred to above;

(2) the nucleic acid A binds complementary fragments from a solution of nucleic acids which display defective base pairing (mutation fragment); a primer B is also added; the mutation fragment may also be identical with the primer B;

(3) the mutation fragment and the primer B are extended;

(4) the extended molecules are ligated with one another; and

(5) the extended primer B is transferred to a second surface, as described in FIG. 1A, Stage 7.

The advantage of this approach is that the second surface only contains mutated molecules.

FIG. 9 shows a process for cloning and sequencing genomic fragments:

(1) following restriction-digesting the DNA fragments are ligated with two different linkers, which specify the sequence of the primers A and B to be used, the genomic fragments are singled out in accordance with the process shown in FIG. 3A; (In the propagation process only those fragments are amplified which carry the various linkers)

(2) the fragments that are singled out and amplified are sorted by means of hybridization ("chromosome walking");

(3) the sorted fragments are propagated individually, split with the use of restriction endonucleases, and subcloned; and

(4) the subcloned fragments can again be sorted using hybridization techniques.

FIG. 10 shows a process for the functional analysis of genomic fragments.

(1) DNA fragments are sorted, as described in FIG. 9;

(2) the single-stranded fragments are augmented by chemical or enzymatic synthesis to produce double strands; and

(3) the fragments are brought into contact with factors (e.g. repressor proteins, activator proteins). Proof of specific binding with specific fragments of nucleic acids enables a functional allocation to take place, in a genomic context.

FIG. 11 shows a process for the parallel quantifying of the gene expression.

(1) following reverse transcription from mRNA, the DNA fragments are provided with linkers and are singled out, as described in FIG. 9;

(2) the cDNA fragments are sorted;

(3) the cDNA fragments are sequenced;

(4) copies of the sorted and sequenced libraries are brought into contact with cellular mRNAs from a healthy cell. The specific hybridization events are confirmed using known, state-of-the-art processes (e.g. fluorescent reporter groups);

(5) in an analog manner the cellular mRNAs from a pathologically altered cell (e.g. a tumorous cell) are brought into contact with another copy of the library; and

(6) a comparison of gene expression samples that have been quantified in this way enables identification to be made of the gene associated with the disease.

FIG. 12 shows the use of the process as a means of improving the signal-to-noise ratio in the detection process:

(1) a library is set up in accordance with the process described in FIG. 9 or 11;

(2) the library is brought into contact with the single-stranded DNA or RNA to be analyzed;

(3) the hybridizing DNAs or RNAs are transferred to the opposite surface, while retaining the site information;

(4) steps 2 and 3 are repeated; and

(5) step 4 can be carried out or repeated any number of times. Signal measurement can be conducted using sensitive scanning techniques such as scanning electrochemical microscopy (SECM) or atomic force microscopy (AFM).

FIG. 13 shows the process when used for the functional allocation of proteins:

(1) a library is compiled in accordance with the process described in FIG. 9 or FIG. 11, in which one of the two linkers used contains a starter sequence for an RNA polymerase; (The promoter can also be added afterwards by means of the process described in FIG. 7)

(2) the DNA fragments are sorted;

(3) chemical or enzymatic synthesis is used to make the single-stranded fragments into double-strands;

(4) in-vitro transcription is used to translate the double-stranded DNA fragments into RNA, the RNAs thus created are transferred to a new surface by the application of an electrical field and the provision of site information; and

(5) the RNA library is translated into proteins by in-vitro translation, the proteins thus created are transferred to a new surface by the application of an electrical field with the preservation of site information: (Because proteins may possess widely differing net charges, it is preferable to repeat the in-vitro translation stage, carrying out the transfer stage with reversed polarity) The protein library is brought into contact with one or more factors (proteins, RNAs, DNAs, other molecules with a biological or chemical origin). The specific binding events are demonstrated by known, state-of-the-art processes. Proof of the specific binding permits the simultaneous detection of functional interactions.

FIG. 14 shows how the process is used to screen combinatory protein libraries:

(1) a library of oligonucleotides is produced by chemical synthesis in accordance with known, state-of-the-art processes;

(2) the library of oligonucleotides is extended in the direction of the 3'- and 5'-terminals by means of the process described in FIG. 7; (The extending sequences provide codes, for example for the constant regions of a single chain antibody)

(3) chemical or enzymatic synthesis is used to make the single-stranded fragments into double-strands;

(4) in-vitro transcription is used to translate the double-stranded DNA fragments into RNA, the RNAs thus created are transferred to a new surface by the application of an electrical field and the preservation of site information;

(5) the RNA library is translated into proteins by in-vitro translation, the proteins thus created are transferred to a new surface by the application of an electrical field and the provision of site information; and

(6) the protein library is brought into contact with one or more factors (proteins, RNAs, DNAs, other molecules with a biological or chemical origin). The specific binding events are demonstrated by known, state-of-the-art processes. Proof of the specific binding permits the simultaneous detection of functional interactions.

Various features and advantages of the present invention are further illustrated by the following non-limiting example.

EXAMPLE I

The invention is explained using the following working example, i.e. a process for propagating nucleic acids on two surfaces: Synthesis of the primer A led to a biotin label at its 5' end, whereas the primer B received a fluorescein label at its 5' end. The labels were produced in accordance with the usual state-of-the-art methods, for example phosphoamide chemistry. Primer A was coupled with a membrane A. A

paper membrane was used as the surface, since this is known to allow the permeation of nucleic acids.

Covalent, coupled streptavidin was applied to this membrane. The coupling took place through the reciprocal action between the biotin and streptavidin. A DNA strand that is complementary to primer A was hybridized on this primer A. The primer A was extended at the 3' end with taq-polymerase or with the Klenow fragment of the polymerase 1. Subsequent denaturing was carried out by heating to 90° C. (alternatively denaturing can also be carried out using a common denaturing reagent). To hybridize the primer B the denaturing solution was replaced by a solution containing TRIS-borate-EDTA-buffers. The primer B was then also extended in the same way as primer A (see also FIGS. 3 and 4).

Using standard electrophoresis apparatus the membrane A was placed on a gel, PAGE gel being preferred. Previously a membrane B was applied to the reverse of the gel, on which fluorescein antibodies had been immobilized. This "sandwich" was held mechanically in a frame, where it was stabilized, and then placed in an electrophoresis chamber designed to ensure separation of both buffers, i.e. the sandwich separates the anode and cathode compartments. A voltage was applied to match the thickness of the gel (in this case 300 v), and subsequently a denaturing agent was flushed into the electrolyte in the (−)-compartment (e.g. a solution of urea). After being heated to 70° C. electrophoresis was carried out. This caused the extended primer B to become detached and travel through the gel layer on the membrane B, where it was subsequently bound and immobilized by binding onto the fluorescein antibodies.

The sandwich was then removed from the electrophoresis chamber, the membranes were removed from the gel, and in accordance with the method described previously, the primer A was again hybridized and extended. Both membranes were then laid on a fresh PAGE gel shaped to enable both membranes to fit accurately in their original orientation, retaining the site information. Then, as described above, electrophoresis was carried out, but in such a way that the membrane B, previously in the (+) compartment (anode compartment), was now allocated to the (−) compartment (cathode compartment).

Alternatively the experiment can be carried out in a microfluid apparatus, in which the membranes and electrodes are firmly positioned, whereas the anode and cathode compartments are flushed out separately and can be flushed using the corresponding reagent. The weakly cross-linked gels that are commonly used in capillary electrophoresis are used in this apparatus. The gels are then replaced for each process.

Alternatively the experiment can also be carried out using activatable reactive primers. In the context of this invention, activatable reactive primers are understood as those possessing a reactive function, and whose reactivity can be influenced by the choice of suitable external conditions. These external conditions may be of a chemical, electrochemical or photochemical nature. An oligonucleotide possessing a cysteine unit via an amino linker whose thiol group is protected in the form of a 2-thiopyrideisulphide group is one example of an activatable reactive primer. In this case the membrane contains carboxy groups in the form of reactive thioesters. Redox-neutral reaction conditions are used in hybridization. After traveling through the field the extended primer reaches a membrane on or in which reductive conditions apply. Reductive conditions are created by the presence, for example, of thiols such as dithioerytrol or dithiothreitol. Disulphide replacement reactions occur in the presence of these reagents, leading to the splitting off of 2-thiopyridons, with the result that the thiol groups released in this way can react on the extended primer with the thioester on the membrane. This reaction initially leads to the formation of a thioester which, due to the presence of the adjacent intramolecular amino group of the cysteine, reacts to form an amide.

One feature of the invention resides broadly in a method for propagating ligands and receptors on at least two surfaces, encompassing one or more of the following cycles: a) Immobilization of a ligand on a first surface of a solid phase; b) Adding a solution of receptors and binding complementary receptors to the ligands; c) Transferring the receptor to an additional surface and immobilizing the receptor at that location; d) Attaching an additional ligand to the immobilized receptor; e) Transferring the ligand to a surface and immobilizing it at that location.

Another feature of the invention resides broadly in a method according to Claim 1, in which the surface in stage (c) and (e) is a second surface which is spatially separated from the first.

Yet another feature of the invention resides broadly in a method in which the transfer in stage (c) and (e) is achieved by the application of an electrical field.

Still another feature of the invention resides broadly in a method in which the electrical field is applied between the first and second surface.

A further feature of the invention resides broadly in a method for the enzymatic propagation of nucleic acids on at least two surfaces, encompassing one or more of the following amplification cycles: a) Immobilization of a first primer on one of the first surfaces of a solid phase; b) Administration of a solution of nucleic acids and binding of complementary fragments to the first primer; c) Extension of the first primer at its 3' end, corresponding to the complementary fragment, by means of a polymerase; d) Release of the complementary fragments; e) Attaching a second primer to the 3' end of the extended nucleic acid; f) Extension of the second primer at its 3' end by means of a polymerase; g) Transfer of the second primer to another surface and immobilization of the extended primer; h) Attaching another first primer to the 3' end of the second, extended primer; Another feature of the invention resides broadly in a method in which the surface in stage (g) is a second surface which is spatially separated from the first.

Yet another feature of the invention resides broadly in a method in which the transfer in stage (g) is achieved by the application of an electrical field.

Still another feature of the invention resides broadly in a method in which the electrical field is applied between the first and second surface.

A further feature of the invention resides broadly in a method in which the following amplification stages occur following the transfer to the second surface: a) Extension of this first primer to its 3' end, corresponding to the complementary fragment, by means of a polymerase; b) Transfer of the extended primer to the first or another surface and immobilization of the primer; c) Attachment of another second primer to the 3' end of the extended first primer.

Another feature of the invention resides broadly in a method for copying nucleic acids from a first to a second surface, encompassing the following stages of the method: a) Immobilization of nucleic acids through a reaction with the surface of a solid carrier; b) Production of double-stranded molecules by hybridization of complementary

single strands; or c) chemical or enzymatic ligation of complementary fragments; d) chemical or enzymatic extension of complementary primers; e) Transfer of complementary strands to a second surface, and their immobilization.

Yet another feature of the invention resides broadly in a method in which the transfer in stage (c) occurs by the application of an electrical field, in which the electrical field is applied between the first and second surface.

Still another feature of the invention resides broadly in a method in which the nucleic acids on the solid carrier are arranged two-dimensionally and are transferred in this order, while retaining site information.

A further feature of the invention resides broadly in a method in which the solid phase material is selected from organic or anorganic material or from a hybrid of these materials, and represents a two- or three-dimensional matrix.

Another feature of the invention resides broadly in a method in which the immobilization takes place through covalent or non-covalent binding.

Yet another feature of the invention resides broadly in a method in which the nucleic acids, ligands, receptors or their derivatives are provided with a detectable label.

Still another feature of the invention resides broadly in a method in which the label is selected from the group of radioisotopes, stable isotopes, enzymes, immunoreactive compounds, fluorescence or luminescence chemicals, chromophores, metals or charged particles.

A further feature of the invention resides broadly in a method according to at least one of the previous claims, in which the solution of nucleic acids includes D- and/or L-nucleic acids.

Yet another feature of the invention resides broadly in a method according to at least one of the previous claims, in which an intermediate layer that can be permeated by nucleic acids and/or ligands/receptors is placed between the surfaces.

Still another feature of the invention resides broadly in a method in which the intermediate layer is selected from the group comprising a gel, a membrane, a polymer, a ceramic and/or a co-called capillary tube array.

A further feature of the invention resides broadly in a method in which the nucleic acids are provided with a positively charged group of headings.

Another feature of the invention resides broadly in a method in which, during the respective transfer stage within the respective method sequence, the scale can be reduced and/or increased, while retaining the site information.

Yet another feature of the invention resides broadly in the use of the method according to Claims 5 to 9 for cloning genomic fragments of DNA, cDNA und RNA.

Still another feature of the invention resides broadly in the use for subcloning following restriction-digesting.

A further feature of the invention resides broadly in the use of the method for strengthening an immunological ligand/receptor pair.

Another feature of the invention resides broadly in the use of the method for strengthening the ligand signal.

Yet another feature of the invention resides broadly in the use of the method for sorting adjacent fragments by using hybridization techniques (chromosome walking).

Still another feature of the invention resides broadly in the use of the method for the copying of gene chips.

The components disclosed in the various publications, disclosed or incorporated by reference herein, may be used in the embodiments of the present invention, as well as equivalents thereof.

The appended drawings in their entirety, including all dimensions, proportions and/or shapes in at least one embodiment of the invention, are accurate and are hereby included by reference into this specification.

All, or substantially all, of the components and methods of the various embodiments may be used with at least one embodiment or all of the embodiments, if more than one embodiment is described herein.

All of the patents, patent applications and publications recited herein, and in the Declaration attached hereto, are hereby incorporated by reference as if set forth in their entirety herein.

The corresponding foreign and international patent publication applications, namely, Federal Republic of Germany Patent Application No. 198 54 946.6, filed on Nov. 27, 1998, entitled KOPIEREN UND KLONIEREN AN OBERFLÄCHEN, having inventors Jens Peter FÜRSTE, Sven KLUSSMANN, and Thomas KLEIN, and DE-OS 198 54 946.6, having inventors Jens Peter FÜRSTE, Sven KLUSSMANN, and Thomas KLEIN, and DE-PS 198 54 946.6, having inventors Jens Peter FÜRSTE, Sven KLUSSMANN, and Thomas KLEIN, and International Application No. PCT/DE99/03856, entitled KOPIEREN UND KLONIEREN AN OBERFLÄCHEN, filed on Nov. 26, 1999, having inventors Jens Peter FÜRSTE, Sven KLUSSMANN, and Thomas KLEIN, as well as their published equivalents, and other equivalents or corresponding applications, if any, in corresponding cases in the Federal Republic of Germany and elsewhere, and the references and documents cited in any of the documents cited herein, such as the patents, patent applications and publications, are hereby incorporated by reference as if set forth in their entirety herein. All the patents, patent applications and publications anywhere in the present application, such as the references and documents cited in any of the documents cited herein, are hereby incorporated by reference as if set forth in their entirety herein.

The details in the patents, patent applications and publications may be considered to be incorporable, at applicant's option, into the claims during prosecution as further limitations in the claims to patentably distinguish any amended claims from any applied prior art.

The invention as described hereinabove in the context of the preferred embodiments is not to be taken as limited to all of the provided details thereof, since modifications and variations thereof may be made without departing from the spirit and scope of the invention.

What is claimed is:

1. A method for propagating ligands and receptors on at least two surfaces, comprising one or more of the following cycles:

(a) immobilizing a first ligand on a first surface of a solid phase;

(b) adding a solution of receptors and binding complementary receptors to the first ligand;

(c) transferring the receptor to a location on a second surface and immobilizing the receptor at that location;

(d) attaching an additional ligand to the immobilized receptor; and

(e) transferring the additional ligand to the first surface and immobilizing it at that location, wherein the steps set forth above may be repeated multiple times.

2. The method according to claim 1, wherein the transfer in stage (c) and (e) is achieved by the application of an electrical field.

3. The method according to claim 2, wherein the electrical field is applied between the first and second surface.

4. The method according to claim 1, wherein the solid phase is selected from a member of the group consisting of organic or inorganic material or from a hybrid of these materials, and represents a two- or three-dimensional matrix.

5. A method for strengthening an immunological ligand/receptor pair, comprising one or more of the following cycles:

(a) immobilizing a first ligand on a first surface of a solid phase;

(b) adding a solution of receptors and binding complementary receptors to the first ligand;

(c) transferring the receptor to a location on a second surface and immobilizing the receptor at that location;

(d) attaching an additional ligand to the immobilized receptor; and

(e) transferring the additional ligand to the first surface and immobilizing it at that location, thus strengthening an immunological ligand/receptor pair, wherein the steps set forth above may be repeated multiple times.

6. A method for strengthening a ligand signal, comprising one or more of the following cycles:

(a) immobilizing a first ligand on a first surface of a solid phase;

(b) adding a solution of receptors and binding complementary receptors to the first ligand;

(c) transferring the receptor to a location on a second surface and immobilizing the receptor at that location;

(d) attaching an additional ligand to the immobilized receptor; and

(e) transferring the additional ligand to the first surface and immobilizing it at that location, thus strengthening an ligand/signal, wherein the steps set forth above may be repeated multiple times.

7. A method for the enzymatic propagation of a nucleic acid sequence on at least two surfaces, comprising:

(a) immobilizing a first primer on at least one first surface of a solid phase;

(b) administering a solution of nucleic acids comprising complementary fragments to the first primer;

(c) binding of complementary fragments to the first primer;

(d) extending the first primer at its 3' end, corresponding to the complementary fragment by means of a polymerase;

(e) releasing the complementary fragments;

(f) attaching a second primer to the 3' end of the extended nucleic acid;

(g) extending the second primer at its 3' end by means of a polymerase;

(h) transferring the second primer to another surface and immobilization of the extended primer; and

(i) attaching another first primer to the 3' end of the second extended primer for further extending of the first primer.

8. The method according to claim 7, wherein the surface in step (h) is a second surface which is spatially separated from the first.

9. The method according to claim 7, wherein the transfer in step (h) is achieved by the application of an electrical field.

10. The method according to claim 9, wherein the electrical field is applied between the first and second surface.

11. The method according to claims 7, further comprising the following amplification steps

(j) extending this first primer to its 3' end, corresponding to the complementary fragment, by means of a polymerase;

(k) transferring of the extended primer to the first or another surface and immobilization of the extended primer thereon; and

(l) attaching of another second primer to the 3' end of the extended first primer.

12. A method for cloning genomic fragments of DNA, cDNA and RNA, comprising:

(a) immobilizing a first primer on at least one first surface of a solid phase;

(b) administering a solution of nucleic acids comprising complementary fragments to the first primer;

(c) binding of complementary fragments to the first primer;

(d) extending the first primer at its 3' end, corresponding to the complementary fragment by means of a polymerase;

(e) releasing the complementary fragments;

(f) attaching a second primer to the 3' end of the extended nucleic acid;

(g) extending the second primer at its 3' end by means of a polymerase;

(h) transferring the second primer to another surface and immobilization of the extended primer;

(i) attaching another first primer to the 3' end of the second extended primer for further extending of the first primer;

(j) extending this first primer to its 3' end, corresponding to the complementary fragment, by means of a polymerase;

(k) transferring of the extended primer to the first or another surface and immobilization of the extended primer and (m) cloning fragments produced by the above steps thereon;

(l) attaching of another second primer to the 3' end of the extended first primer.

13. The method according to claim 11, wherein the nucleic acids on the solid phase surface are arranged two-dimensionally and are transferred in this order, while retaining site information.

14. The method according to claim 11, wherein the solid phase surface is selected from a member of the group consisting of organic or inorganic material or from a hybrid of these materials, and represents a two- or three-dimensional matrix.

15. The method according to claims 11, wherein immobilization on the solid phase takes place through covalent or non-covalent binding.

16. The method according to claim 7, where the nucleic acids are provided with a detectable label.

17. The method according to claim 16, wherein the label is selected from a member of the group consisting of radioisotopes, stable isotopes, enzymes, immunoreactive compounds, fluorescence or luminescence chemicals, chromophores, metals or charged particles.

18. The method according to claim 7, wherein the nucleic acids include D- and/or L-nucleic acids.

19. The method according to claim 7 wherein an intermediate layer is placed between the surfaces.

20. The method according to claim 19, wherein the intermediate layer is selected from the group consisting of gel, a membrane, a polymer, a ceramic and capillary tube array.

21. The method according to claim 7, wherein the nucleic acids are each provided with a positive charge.

22. A method for sorting adjacent fragments by using hybridization techniques, comprising:

(a) immobilizing a first primer on at least one first surface of a solid phase;

(b) administering a solution of nucleic acids comprising complementary fragments to the first primer;

(c) binding of complementary fragments to the first primer;

(d) extending the first primer at its 3' end, corresponding to the complementary fragment by means of a polymerase;

(e) releasing the complementary fragments;

(f) attaching a second primer to the 3' end of the extended nucleic acid;

(g) extending the second primer at its 3' end by means of a polymerase;

(h) transferring the second primer to another surface and immobilization of the extended primer;

(i) attaching another first primer to the 3' end of the second extended primer for further extending of the first primer;

(j) extending this first primer to its 3' end, corresponding to the complementary fragment, by means of a polymerase;

(k) transferring of the extended primer to the first or another surface and immobilization of the extended primer thereon;

(l) attaching of another second primer and (m) sorting adjacent fragments produced by the above steps to the 3' end of the extended first primer.

* * * * *

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 6,534,271 B2
DATED : March 18, 2003
INVENTOR(S) : Jens Peter Fürste et al.

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

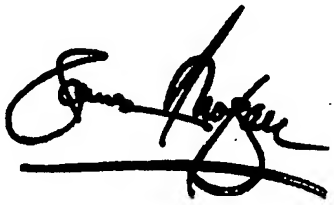
Column 25,
Line 38, before "ligand", delete "an" and insert -- a --.

Column 26,
Lines 41 and 42, after "primer", delete "and (m) cloning fragments produced by the above steps thereon;" and insert -- thereon; --.
Line 44, after "primer", delete "." and insert -- ; and (m) cloning fragments produce by the above steps. --.

Column 28,
Lines 16 and 17, after "primer", delete "and (m) sorting adjacent fragments produced by the above steps".
Line 18, after "primer", delete "." and insert -- ; and (m) sorting adjacent fragments produce by the above steps. --.

Signed and Sealed this

Twenty-sixth Day of August, 2003



JAMES E. ROGAN
Director of the United States Patent and Trademark Office

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 6,534,271 B2
DATED : March 18, 2003
INVENTOR(S) : Jens Peter Fürste et al.

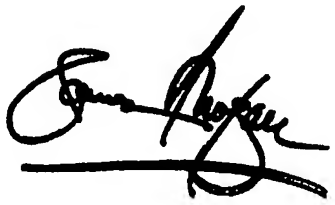
It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 26,
Line 44, after "fragments", delete "produce" and insert -- produced --.

Column 28,
Line 18, after "fragments", delete "produce" and insert -- produced --.

Signed and Sealed this

Sixteenth Day of December, 2003



JAMES E. ROGAN
Director of the United States Patent and Trademark Office

US006632641B1

(12) **United States Patent**
Brennan et al.

(10) **Patent No.: US 6,632,641 B1**
(45) **Date of Patent: *Oct. 14, 2003**

(54) **METHOD AND APPARATUS FOR PERFORMING LARGE NUMBERS OF REACTIONS USING ARRAY ASSEMBLY WITH RELEASABLE PRIMERS**

(75) Inventors: **Thomas M. Brennan**, San Francisco, CA (US); **Francois Chatelain**, San Francisco, CA (US); **Mark Berninger**, North Potomac, MD (US)

(73) Assignee: **Metrigen, Inc.**, Palo Alto, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

This patent is subject to a terminal disclaimer.

(21) Appl. No.: **09/686,597**

(22) Filed: **Oct. 10, 2000**

Related U.S. Application Data

(63) Continuation of application No. 09/684,736, filed on Oct. 6, 2000.
(60) Provisional application No. 60/158,315, filed on Oct. 8, 1999.

(51) **Int. Cl.**[7] **C12Q 1/68**; C12P 19/34; C12M 1/34; C07H 21/02; C07H 21/04

(52) **U.S. Cl.** **435/91.2**; 435/6; 435/7.1; 435/91.1; 435/287.2; 530/22.1; 530/23.1; 530/24.3; 530/24.31; 530/24.32; 530/24.33

(58) **Field of Search** 435/6, 7.1, 91.1, 435/91.2; 536/22.1, 23.1, 24.3–24.33

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,683,202	A	7/1987	Mullis	435/91
4,834,946	A	5/1989	Levin	422/101
5,143,854	A	9/1992	Pirrung et al.	436/518
5,202,231	A	4/1993	Dramanac et al.	435/6
5,412,087	A	5/1995	McGall et al.	536/24.3
5,445,934	A	8/1995	Fodor et al.	435/6
5,445,943	A	8/1995	Hoenes	435/26
5,474,796	A	12/1995	Brennan	427/2.13
5,489,678	A	2/1996	Fodor et al.	536/22.1
5,492,806	A	2/1996	Drmanac et al.	435/5
5,525,464	A	6/1996	Drmanac	435/6

(List continued on next page.)

FOREIGN PATENT DOCUMENTS

EP	717113 A2	6/1996
WO	WO 92/15712	9/1992

(List continued on next page.)

OTHER PUBLICATIONS

Abramson et al., "Nucleic acid amplification technologies," *Curr. Opin. Biotechnol. 4*:41–47 (1993).
Abravaya et al., "Detection of point mutations with a modified ligase chain reaction", *Nucleic Acids Res. 23*:675–682 (1995).
Adinolfi et al., "Solid Phase Synthesis of Oligosaccharides," *Tetrahedron Lett. 37*(28):5007–5010 (1996).

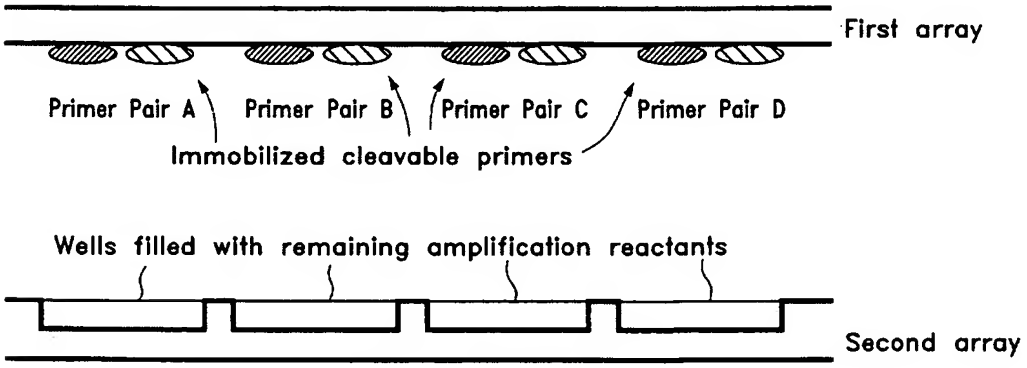
(List continued on next page.)

Primary Examiner—Jeffrey Siew
(74) *Attorney, Agent, or Firm*—Howrey, Simon, Arnold & White, LLP; Albert P. Halluin

(57) **ABSTRACT**

The present invention relates to a method and apparatus for performing a large number of reactions using array assembly. In particular, the present invention features a method and apparatus for performing a large number of chemical and biological reactions by bringing two arrays into close apposition and allowing reactants on the surfaces of two arrays to come into contact. The present invention is exemplified by performing a large number of polynucleotide amplification reactions using array assembly. In addition, the present invention features a method and apparatus for coupling the amplification of polynucleotides and the detection of sequence variations, expression levels, and functions thereof.

25 Claims, 30 Drawing Sheets



U.S. PATENT DOCUMENTS

5,545,531	A	*	8/1996	Rava et al.
5,545,568	A		8/1996	Ellman 436/518
5,556,749	A		9/1996	Mitsuhashi et al. 435/6
5,571,639	A		11/1996	Hubbell et al. 430/5
5,614,608	A		3/1997	Krchnak et al. 530/334
5,650,277	A		7/1997	Navot et al. 435/6
5,667,972	A		9/1997	Drmanac et al. 435/6
5,679,773	A		10/1997	Holmes 530/334
5,691,141	A		11/1997	Köster 435/6
5,695,940	A		12/1997	Drmanac et al. 435/6
5,700,637	A		12/1997	Southern 435/6
5,700,642	A		12/1997	Monforte et al. 435/6
5,710,028	A		1/1998	Eyal et al. 435/91.1
5,739,386	A		4/1998	Holmes 562/437
5,744,305	A		4/1998	Fodor et al. 435/6
5,800,992	A		9/1998	Fodor et al. 435/6
5,830,655	A		11/1998	Monforte et al. 435/6
5,837,832	A		11/1998	Chee et al. 536/22.1
5,846,943	A		12/1998	Hindsgaul et al. 514/25
5,858,653	A		1/1999	Duran et al. 435/6
5,858,659	A		1/1999	Sapolsky et al. 435/6
5,871,928	A		2/1999	Fodor et al. 435/6
5,888,819	A		3/1999	Goelet et al. 435/5
5,889,165	A		3/1999	Fodor et al. 536/22.1
5,917,016	A		6/1999	Holmes 530/334
5,919,626	A		7/1999	Shi et al. 435/6
5,922,534	A		7/1999	Lichtenwalter 435/6
5,927,547	A		7/1999	Papen et al. 222/57
5,929,208	A		7/1999	Heller et al. 530/333
5,972,619	A		10/1999	Drmanac et al. 435/6
5,985,551	A		11/1999	Brennan 435/6
5,985,557	A		11/1999	Prudent et al. 435/6
5,985,761	A		11/1999	Saprks et al. 438/669
6,001,567	A		12/1999	Brow et al. 435/6
6,017,696	A	*	1/2000	Heller
6,018,041	A		1/2000	Drmanac et al. 536/24.3
6,025,136	A		2/2000	Drmanac 435/6
6,028,189	A		2/2000	Blanchard 536/25.3
6,030,782	A		2/2000	Anderson et al. 435/6
6,040,138	A		3/2000	Lockhart et al. 435/6
6,043,031	A		3/2000	Köster et al. 435/6
6,054,270	A		4/2000	Southern 435/6
6,074,823	A		6/2000	Köster 435/6
6,083,763	A		7/2000	Balch 436/518
6,090,995	A		7/2000	Reich et al. 623/11
6,103,479	A		8/2000	Taylor 435/7.2
6,197,506	B1		3/2001	Fodor et al. 435/6
6,210,894	B1		4/2001	Brennan 435/6
6,218,118	B1	*	4/2001	Sampson et al.
6,238,869	B1	*	5/2001	Kris et al. 435/6
6,288,220	B1	*	9/2001	Kambara et al.
6,291,183	B1		9/2001	Pirrung et al. 435/6
6,300,066	B1	*	10/2001	Gray et al.
6,300,070	B1	*	10/2001	Boles et al.
6,309,822	B1		10/2001	Fodor et al. 435/6
6,309,823	B1		10/2001	Cronin et al. 435/6
6,309,831	B1		10/2001	Goldberg et al. 435/6
6,310,189	B1		10/2001	Fodor et al. 435/6
6,322,968	B1	*	11/2001	Head et al.

FOREIGN PATENT DOCUMENTS

WO	WO 93/09250	5/1993
WO	WO 93/17126	9/1993
WO	WO 93/17136	9/1993
WO	WO 94/11530	5/1994
WO	WO 95/11995	5/1995
WO	WO 97/28282	8/1997
WO	WO 97/35198	9/1997
WO	WO 97/43447	11/1997
WO	WO 97/45730	12/1997
WO	WO 98/09735	3/1998
WO	WO 98/21221	5/1998
WO	WO 98/22487	5/1998
WO	WO 98/28438	7/1998
WO	WO 98/30883	7/1998
WO	WO 98/33586	8/1998
WO	WO 98/38490	9/1998
WO	WO 98/38846	9/1998
WO	WO 98/41531	9/1998
WO	WO 98/46247	10/1998
WO	WO 98 47003	10/1998
WO	WO 98/50403	11/1998
WO	WO 98/54362	12/1998
WO	WO 98/56954	12/1998
WO	WO 99/05308	2/1999
WO	WO 99/06593	2/1999
WO	WO 99/06834	2/1999
WO	WO 99/07888	2/1999
WO	WO 99/09073	2/1999
WO	WO 99/14228	3/1999
WO	WO 99/21957	5/1999
WO	WO 99/27137	6/1999
WO	WO 99/37812	7/1999
WO	WO 99/39004	8/1999
WO	WO 99/47701	9/1999
WO	WO 99/54509	10/1999
WO	WO 99/58708	11/1999
WO	WO 00/03246	1/2000
WO	WO 00/17624	3/2000
WO	WO 00/17643	3/2000
WO	WO 00/50872	8/2000

OTHER PUBLICATIONS

Albericio et al., "Convergent Solid-Phase Peptide Synthesis," *Methods Enzymol. 289*:313–316 (1997).

Andres, et al., "Transition-metal-mediated reactions in combinatorial synthesis," *Curr. Opin. Chem. Biol. 2*:353–362 (1998).

Atherton et al., *Solid Phase Peptide Synthesis: A practical approach*, IRL press, London (1989).

Ausubel, et al., *Current Protocols in Molecular Biology, vol. 1–2, John Wiley & Sons (1989)*.

Beier et al., "Versatile Derivatisation of solid support media for covalent bonding on DNA-microchips," *Nucleic Acids Res. 27*(9):1970–1977 (1999).

Blanchard, et al., "Synthetic DNA Arrays", *Biosensors and Bioelectronics 11*:687–690 (1996).

Blixt et al., "Solid-Phase Enzymatic Synthesis of a Sialyl Lewis X Tetrasaccharide on a Sepharose Matrix," *J. Org. Chem. 63*:2705–2710 (1998).

Brzoska, et al., "Evidence of a transition temperature for the optimum deposition of grafted monolayer coatings," *Nature 360*:719–721 (1992).

Buhr, et al. "Oligodeoxynucleotides containing C-7 propyne analogs of 7-deaza-2'-deoxyguanosine and 7-deaza-2'-deoxyadenosine," *Nucleic Acids Res. 24*(15):2974–2980 (1996).

Bulyk et al. "Quantifying DNA-protein interations by double-stranded DNA arrays," *Nature Biotechnology, 17*:573–577 (1999).

Burg et al., "Real-Time Fluorescense Detection of RNA Amplified by Qβ Replicase," *Anal. Biochem. 230*:263–272 (1995).

Cantor and Schimmel, "Part 1: The conformation of Biological macromolecules," Biophycical Chemistry, San Francisco, W.H. Freeman (1980).

Cantor and Smith, *Genomics: the science and technology behind the human genome project*, John Wiley & Sons (1999).

Case-Green, et al. "Analyzing genetic information with DNA arrays," *Cur. Opin. In Chem. Biol.* 2:404–410 (1998).

Czarnik, et al., "Guest Editorial," *Accounts Chem. Rev.* 29:112–170 (1996).

Danishefsky et al., "A Strategy for the Solid-Phase Synthesis of Oligosaccharides," *Science* 260:1307–1309 (1993).

DeRisi, et al., "Exploring the Metabolic and Genetic Control of Gene Expression on A Genomic Scale," *Science* 278:680–686 (1997).

de Wildt, et al., "Antibody arrays for high-throughtput screening of antibody-antigen interactions," *Nature Biotechnol.* 18:989 (2000).

Drmanac et al., "Sequencing of Megabase Plus DNA by Hybridization: Theory of the Method," *Genomics* 4:114–28 (1989).

Drmanac et al., "Accurate sequencing by hybridization for DNA diagnostics and individual genomics," *Nature Biotechnology* 16:54–58 (1998).

Duggan, et al., "Expression profiling using cDNA microarrays," *Nature Genetics Supplement* 21:10–14 (1999).

Eckert et al., "DNA Polymerase Fidelity and the Polymerase Chain Reaction," *PCR Methods and Applications* 1:17 (1991).

Edman, et al., "Electric field directed nucleic acid hybridization on microchips," *Nucleic Acids Research*, 25(24):4907–4914 (1997).

Eisenberg, et al., "Protein function in the post-genomic era," *Nature* 405:823–826 (2000).

Fodor, et al. "Light-Directed, Spatially Addressable Parallel Chemical Synthesis," *Science* 251:767–773 (1991).

Fruchtel, "Organic Chemistry and Solid Supports," *Angew. Chem. Int. Ed. Engl.* 35:17–42 (1996).

Gelfand, et al., "ASDB: database of alternatively spliced genes," *Nucleic Acids Res.* 27(1):301–302 (1999).

Gerhold, et al., "DNA chips: promising toys have become powerful tools," TIBS, 24:168–173 (1999).

Gibson, et al. "A Novel Method for Real Time Quantitative RT-PCR," *Genome Res.* 6:995–1001 (1996).

Giesen, et al., "A formula for thermal stability (T_m) prediction of PNA/DNA duplexes," *Nucleic Acids Research* 26(21):5004–5006 (1998).

Good, et al., "Antisense inhibition of gene expression in bacteria by PNA targeted to mRNA," *Nature Biotechnology* 16:355–358 (1998).

Gordon et al. (eds.) Combinatorial Chemistry and Molecular Diversity in Drug Discovery, John Wiley & Son, New York (1997).

Gordon et al., "Applications of Combinatorial Technologies to Drug Discovery. 2. Combinatorial Organic Synthesis, Library Screening Strategies, and Future Directions," *J. Med. Chem.* 37:1385–1401 (1994).

Grant, et al., "Human acetyltransferase polymorphisms," *Mut. Res.* 376:61–70 (1997).

Greenberg et al., "Cleavage of Oligonucleotides from Solid-Phase Supports Using o-Nitrobenzyl Photochemistry," J. *of Org. Chem.* 59:746–753 (1994).

Greenberg, "Photochemical Release of Protected Oliucleotides Containing 3'-Glycolate Termini," *Tetrahedron* 51:29–38 (1995).

Greenberg, "Photochemical Cleavage of Oligonucleotides From Solid Phase Supports," *Tetrahedron Lett.* 34:251–254 (1993).

Gururaja et al., "Solid-Phase Synthesis of Human Salivary Mucin-Derived O-linked Glycopeptide," *Lett Pept. Sci.* 3:79–88 (1996).

Guschin, et al., "Manual Manufacturing of Oligonucleotide, DNA, and Protein Microchips," *Anal. Biochem.* 250:203–211 (1997).

Hammer, et al., "Practical Approach to Solid-Phase Synthesis of C-terminal Peptide Amides under Mild Conditions Based on a Photolysable Anchoring Linkage," *J. Peptide Protein Res.* 36:31–45 (1990).

Heckel, et al., "Oligosaccharide Synthesis on Controlled-Pore Glass as Solid Phase Material," *Synlett* 171–173 (1998).

Heid, et al., "Real Time Quantitative PCR," *Genome Res.* 6:986–994 (1996).

Hermkens, et al., "Solid-Phase Organic Reactions: A review of the Recent Literature," *Tetrahedron* 52:4527–4554 (1996).

Higuchi, et al. Simultaneous Amplification and Detection of Specific DNA Sequences, *BioTechnology* 10:413–417 (1992).

Higuchi, et al., "Kinetic PCR Analysis: Real-time Monitoring of DNA Amplification Reaction," *BioTechnology* 11:1026–1030 (1993).

Holmes et al., "Reagents for Combinatorial Organic Synthesis: Development of a New o-Nitrobenzyl Photoabile Linker for Solid Phase Synthesis," J. *of Org. Chem.* 60:2318–2319 (1995).

Holmes, et al., "Model Studies for New o-Nitrobenzyl Photolabile Linkers: Substituent Effects on the Rates of Photochemical Cleavage," *J. of Org. Chem.* 62:2370–2380 (1997).

Hughes, et al., "Functional Discovery via a Compendium of Expression Profiles," *Cell* 102:109–126 (2000).

Hyndman, et al., "Software to Determine Optimal Oligonucleotide Sequnces Based on Hybridization Simulation Data," *BioTechniques* 20(6):1090–1097 (1996).

Ishiguro, et al. "Homogeneous Quantitative Assay of Hepatitis C Virus RNA by Polymerase Chain Reaction in the Presence of a Fluorescent Intercalater," *Anal Biochem.* 229:207–213 (1995).

Isaksson and Landegren, "Accessing genomic information: alternatives to PCR," *Curr. Opin. Biotechnol.* 10:11–15 (1999).

Ito et al., "Solid-phase oligosaccharide synthesis and related technologies," *Curr. Opin. Chem. Biol.* 2:701–708 (1998).

Joos et al., "Covalent Attachment of Hybridizable Oligonucleotides to Glass Supports," *Anal. Chem.* 247:96–101 (1997).

Kahl, et al., "High-Yielding method for On-Column Derivatization of Protected Oligodeoxy-nucleotides and Its Application to the Convergent Synthesis of 5',3'-Bis-conjugates," *J. of Org. Chem.* 63:4870–4871 (1998).

Kahl and Greensberg, "Solution-Phase Bioconjugate Synthesis Using Protected Oligonucleotides Containing 3'-Alky Carboxylic Acids," *J. of Org. Chem.*, 64:507–510 (1999).

Kahn, et al., *Modern Methods in Carbohydrate Synthesis*, Harwood Academic, Amsterdam (1996).

Kierzek, et al., "Association of 2'-5'oligoribonucleotides," *Nucleic Acids Research* 20(7): 1865–1690 (1992).

Kihlberg, et al., "Direct Synthesis of Glycosylated Amino Acids from Carbohydrate Peracetates and Fmoc Amino Acids: Solid–Phase Synthesis of Biomedicinally Interesting Glycopeptides," *Methods Enzymol. 289*:221–245 (1997).

Krokan, et al., "DNA glycosylases in the base excision repair of DNA," *Biochem. J. 325*:1–16 (1997).

Kuppuswami, et al., "Single nucleotide primer extension to detect genetic diseases: Experimental application to hemophilia B (factor IX) and cystic fibrosis genes," *Proc. Natl. Acad. Sci. USA 88*:1143–1147 (1991).

Landegren, "The Challengers to PCR: a proliferation of chain reactions," *Curre. Opin. Biotechnol. 7*:95–97 (1996).

Landegren, et al, "Reading Bits of Genetic Inforation: Methods for Single–Nucleotide Polymorphism Analysis," *Genome Research 8*:769–776 (1998).

Lie, et al., "Advances in quantitative PCR technology 5'nuclease assays," *Curr. Opin. In Biotech. 9*:43–48 (1998).

Lin et al., "Ethnic distribution of slow acetylator mutations in the polymorphic N–acetyltransferase (NAT2) gene," *Pharmacogenetics 4*:124–134 (1994).

Lipshutz, et al., "High density synthetic oligonucleotide arrays," *Nature Genetics Supplement 21*:20–24 (1999).

Livak, et al. "Oligonucleotides with Fluorescent Dyes at Opposite Ends Provide a Quenched Probe System Useful for Detecting PCR Product and Nucleic Acid Hybridization," *PCR Methods and Applications 4*:357–362 (1995).

Lloyd–Williams, et al., "Convergent Solid–phase peptide synthesis," *Tetrahedron 49*:11065–11133 (1993).

Lockhart, et al., "Genomics, gene expression and DNA arrays," *Nature 405*:827–836 (2000).

Malek et al., "Nucleic Acid Sequence–Based Amplication (NASBA)," *Methods Mol. Biol. 28*:253–260 (1994).

Masko, et al., "Oligonucleotide hydridisations on glass supports: a novel linker for oligonucleotide synthesis and hydridisation properties of oligonucleotides synthesised in situ," *Nucleic Acids Research 20*(7):1679–1684 (1992).

Mattila et al., "Fidelity of DNA Synthesis by the *Thermococcus litoralis* DNA polymerase–an extremely heat stable enzyme with proofreading activity," *Nucleic Acids Res. 19*(18):4967–4973 (1991).

Mcdevitt et al. "Glycosamino Acids: New Building Blocks for Cominatorial Synthesis," *J. Am. Chem. Soc. 118*:3818–3828 (1996).

McKenzie, et al., "Parallel molecular genetic analysis," *European Journal of Human Genetics 6*:417–429 (1998).

McMinn et al., "Efficient Solution Phase synthesis of Oligonucleotide Conjugates Using Protected Biopolymers Containing 3'–Terminal Alkyl amines," *J. of Org. Chem. 62*:7074–7075 (1997).

McMinn et al., "Novel Solid Phase Synthesis Supports for the Preparation of Oligonucleotides Containing 3'–Alkyl Amines," *Tetrahedron 52*:3827–3840 (1996).

Meldal et al. "Synthetic methods for glycopeptide assembly, and biological analysis of glycopeptide products," *Curr. Opin. Chem. Biol. 1*:552–563 (1997).

Merrifield, "Solid–Phase Synthesis," *Science 232*:342–347 (1986).

Methods Mol. Biol. "Protocols for Oligonucleotides and Analogs" (ed. Sudhir Agrawal) vol. 20 (1984).

Mitsuhashi, "Technical Report: Part 1. Basic Requirements for Designing Optimal Oligonucleotide Probe Sequences," *J. Clinical Laboratory Analysis 10*:277–284 (1996).

Mrksich, et al., "Controlling cell attachment on contoured surfaces with self–assembeld monlayers of alkanethiolates on gold," *Proc. Natl. Acad. Sci. USA 93*:10775–8 (1996).

Mrksich, et al., "Using Self–Assembled Monolayers to understand the interactions of man–made surfaces with proteins and cells," *Ann. Rev. Biophys. Biomol. Struct. 25*:55–78 (1996).

Muller et al., "Self–sustained sequence replication (3SR): An alternative to PCR," *Histochem. Cell Biol. 108*:431–437 (1997).

Nelson, "Rapid Detection of Genetic Mutations Using the Chemiluminescent Hybridization Protection Assay (HPA): Overview Comparision with Other Methods," *Crit. Rev. Clin. Lab. Sci. 35*:369–414 (1998).

Nicolaou et al., "A General and Highly Efficient Solid Phase Synthesis of Oligosaccharides. Total Synthesis of a Heptasaccharide Phytoalexin Elicitor (HPE)," *J. Am. Chem. Soc. 119*:449–450 (1998).

Nielsen, "Applications of peptide nucleic acids," *Current Opinion in Biotechnology 10*:71–75 (1999).

Nguyen, et al., "Modification of DNA duplexes to smooth their thermal stability independently oft heir base content for DNA sequencing by hybridization," *Nucleic Acids Research 25*(15):3059–3065 (1997).

Nguyen, et al., "The stability of duplexes involving AT and/or $G^{4E'}C$ base pairs is not dependent on their AT/$G^{4E'}C$ ratio content. Implication for DNA sequencing by hybridization," *Nucleic Acids Research 26*(18):4249–4258 (1998).

Pandey, et al., "Proteomics to study genes and genomes," *Nature 405*:837–846 (2000).

Paulsen et al., "New solid–phase oligosaccharide synthesis on glycopeptides bound to a solid phase," *J. Chem. Perkin Trans 1*:281–293 (1997).

PCR A Practical Approach (eds. McPherson et al., IRL Press, Oxford, 1991).

PCR2 A Practical Approach (eds. McPherson et al., IRL Press, Oxford, 1995).

PCR Protocols: A Guide to Methods and Applications (eds. Innis, et al.), Academic Press, San Diego, CA, 1990.

PCR Technology: Principles and Applifications for DNA Amplication (ed. H.A. Erlich, Freeman Press, NY, NY, 1992.

Pontius, et al., "Rapid renaturation of complementary DNA strands mediated by cationic detergents: A role for high–probability binding domains in enhancing the dinetics of molecular assembly processes," *Proc. Natl. Acad. Sci. USA 88*:8237–8241 (1991).

Rademann et al., "Repetitive SolidPhase Glycosylation ona an Alkyl Thiol Polymer Leading to Sugar Oligomers Containing 1,2–trans– and 1,2–cis–Glycosidic Linckages," *J. Org. Chem. 62*:3650–3653 (1997).

Rees, et al., "Betaine Can Eliminate the Base Pair Composition Dependence of DNA Melting," *Biochemistry 3*:137–144 (1993).

Rich, et al., "Preparation of a New o–Nitrobenzyl Resin for Solid–Phase Synthesis of tert–Butyloxycarbonyl–Protected Peptide Acids," *J. Am. Chem. Soc. 97*:1575–1579 (1975).

Rich, et al., "Removal of Protected Peptides from an ortho–nitrobenzyl resin by photolysis," *J.C.S. Chem. Comun.* 610–611 (1973).

Roberts, et al., "Signalingand Circuitryof Multiple MAPK Pathways Revealed by a Matrix of Global Gene Expression Profiles," *Science 287*:878–880 (2000).

Rodebaugh et al., "Polymer–Supported Oligosaccharides via n–Pentenyl Glycosides: Methodology for a Carbonhydrate Library," *J. Org. Chem. 62*:5660–5661 (1997).

Rychlik, et al., "A computer program for choosing optimal oligonucleotides for filter hybridization, sequencing and in vitro amplification of DNA," *Nucleic Acids Res. 17*:8543–8551 (1989).

Rychlik, et al., "Optimization of the annealing temperature for DNA amplification in vitro," *Nucleic Acids Res. 18*(21):6409–6412 (1989).

Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

Schena et al., "Microarrays: biotechnology's discovery platform for functional geneomics," *TIBTECH 16*:301–306 (1998).

Schullek et al., "A High–density Screening format for Encoded Combinatiorial Libraries: Asssay Miniaturization and its Application to Enzymatic Reactions," *Anal. Biochem. 246*:20–29 (1997).

Shuster et al., "Solid–Phase Chemical–Enzymatic Synthesis of Glycopeptides and Oligosaccharides," *J. Am. Chem. Soc. 116*:1135–1136 (1994).

Silveira and Orgel, "PCR with detachable primers," *Nucleic Acids Research 23*(6):1083–1084.

Singh–Gasson, et al. "Maskless fabrication of light–directed oligonucleotide microarrays using a digital micromirror array," *Nature Biotechnol. 17*:974–978 (1999).

Singhvi, et al., "Engineering Cell Shapte and Function," *Science, 264*:696–698 (1994).

Sokolov, "Primer extension technique for the detection of single nucleotide in genomic DNA," *Nucleic Acids Res. 18*(12):3671 (1990).

Sosnowski et al., "Rapid determination of singel base mismatch mutations in DAN hybrids by direct electric field control," *Proc. Natl. Acad. Sci. 94*:1119–1123 (1997).

Spielberg, et al., "N–Acetyltransferases: Pharmacogenetics and Clinical Consequences of Polymorphic Drug Metabolism," *J. Pharmacokint. Bipharm. 24*(5):509–519 (1996).

Steward, "Cleavage Methods Following Boc–Based Solid–Phase Peptide Synthesis," *Methods in Enzymol. 289*:29–44 (1997).

Syvänen et al., "A Primer–Guilded Nucleotide Incorporation Assay in the Genetyping of Apolipoprotein E," *Genomics 8*:684–692 (1990).

Syvänen, "From Gels to Chips: 'Minisequencing' Primer Extension for Analysis of Point Mutations and Single Nucleotide Polymorphisms", *Human Mutation 13*:1–10 (1999).

Thompson, et al., "Synthesis and Applications of Small Molecule Libraries," *Chem. Rev. 96*:555–600 (1996).

Toshima et al., "Recent Progess in O–glycosylation Methods and Its application to Natural Products Synthesis," *Chem. Rev. 93*:1503–1531 (1993).

Tyagi, et al., "Multicolor Molecular Beacons for allele discrimination," *Nature Biotechnol. 16*:49–53 (1998).

Uetz, et al. "A comprehensive analysis of protein–protein interactions in Saccharomyces," Nature 403:623 (2000).

Van Ness, et al., "The use of oligodeoxynucleotide probes in chaotrope–based hybridization solutions," *Nucleic Acids Research 19*(19):5143–5151 (1991).

Venkatesan and Greenberg, "Improved Utility of Photolabile Sold Phase Synthesis Supports for the Synthesis of Oligonucleotides Containing 3'–Hydroxyl Termini," *J. of Org. Chem., 61*:525–529 (1996).

Verma and Eckstein, "Modified Oligonucleotides: Synthesis and Strategy for Users," *Annu. Rev. Biochem. 67*:99–134 (1998).

Wagner, et al., "Antisense Gene inhibition by oligonucleotides containing C–5 propyne pyrimidines," *Science 260*:1510–1513 (1993).

Walker, "Empirical Aspects of Strand Displacement Amplication," *PCR Methods Appl. 3*:1–6 (1993).

Wang, "Solid Phase Synthesis of Protected Peptide via Photolytic Clevage of the α–methylphenacyl Ester Anchoring Linkage," *J. Org. Chem. 41*:3258 (1976).

Wang et al., "A New Base–Labile Anchoring Group for Polymer–Supported Oligosaccharide Synthesis," *Chem. Lett.* 273–274 (1995).

Wang, D., et al., "Large–Scale Identification Mapping, and Genotyping of Single–Nucleotide Polymorphisms in the Human Genome," *Science 280*:1077–1082 (1998).

Wetmur, "DNA Probes: Applicationis of the Principles of Nucleic Acid Hybridization," *Critical Reviews in Biochemistry and Molecular Biology 26*:227–259 (1991).

Wiedmann, et al., Ligase Chain Reaction (LCR)—Overview and Applications, *PCR Methods Appl. 3*:S51–64 (1994).

White, "High–Throughput Screening in Drug Metabolism and Pharmacokinetic Support of Drug Discovery," *Annu. Rev. Pharmacol. Toxicol. 40*:133–157 (2000).

Wu, et al., "The Ligation Amplification Reaction (LAR)—Amplification of Specific DNA Sequences Using Sequential Rounds of Template–Dependent Ligation," *Genomics 4*:560–569 (1989).

Yamada, et al., An Efficient Synthesis of Sialoglycoconjugtes on a Peptidase–Sensitive Polymer Support, *Tetrahedron Lett. 36*:9493–9496 (1995).

Yan et al., "Glycosylation on the Merrified Resin Using Anomerica Sulfoxides," *J. Am. Chem. Soc. 116*:6953–6954 (1994).

Yoo et al., "Synthesis of Oligonucleotides Containing 3'Alkyl Carboxylic Acids Using Universal, Photolabile Solid Phase Synthesis Supports," *J. of Org. Chem. 60*:3358–3364 (1995).

Young, "Biomedical Discovery with DNA Arrays," *Cell 102*:9–15 (2000).

Zheng et al., "Solid Support Oligosaccharide Synthesis: Construction of β–Linked Oligosaccharides by Coupling by Glycal Derived Thioethyl Glycosyl Donors," *J. Org. Chem. 63*:1126–1130 (1998).

* cited by examiner

Fig. 1

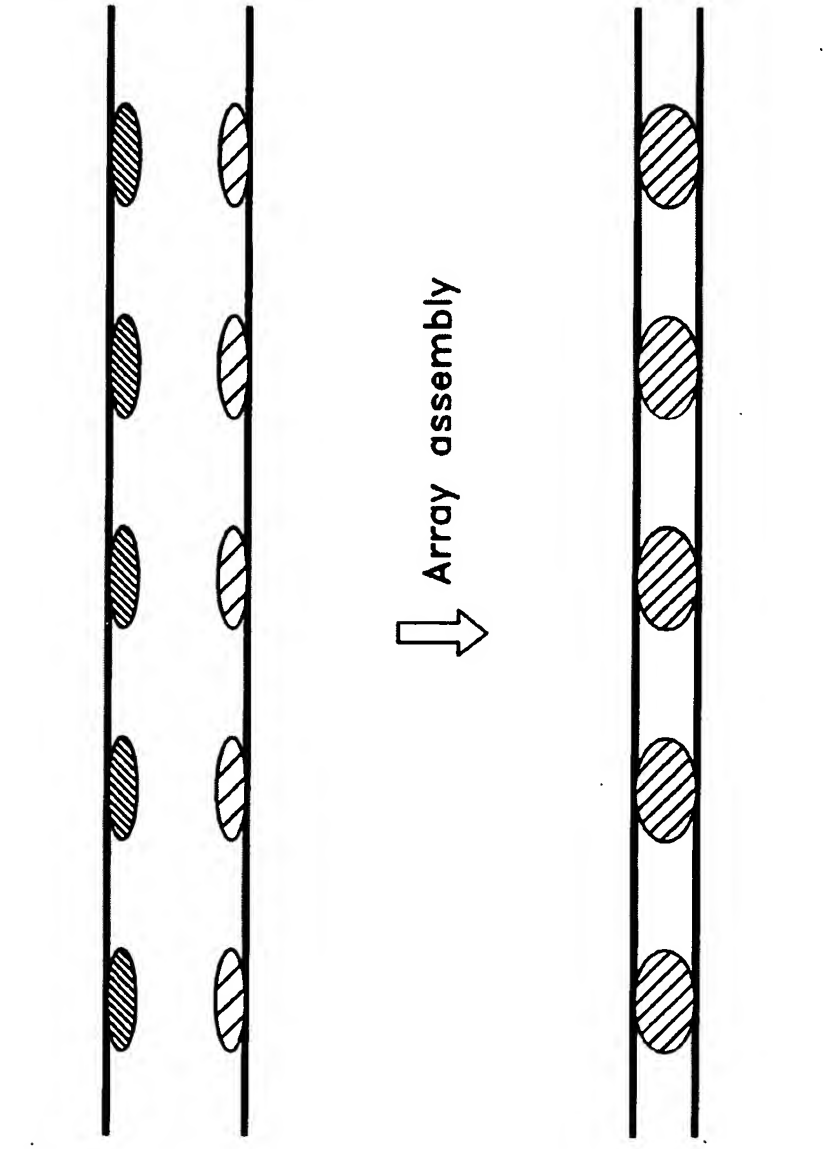


Fig. 2

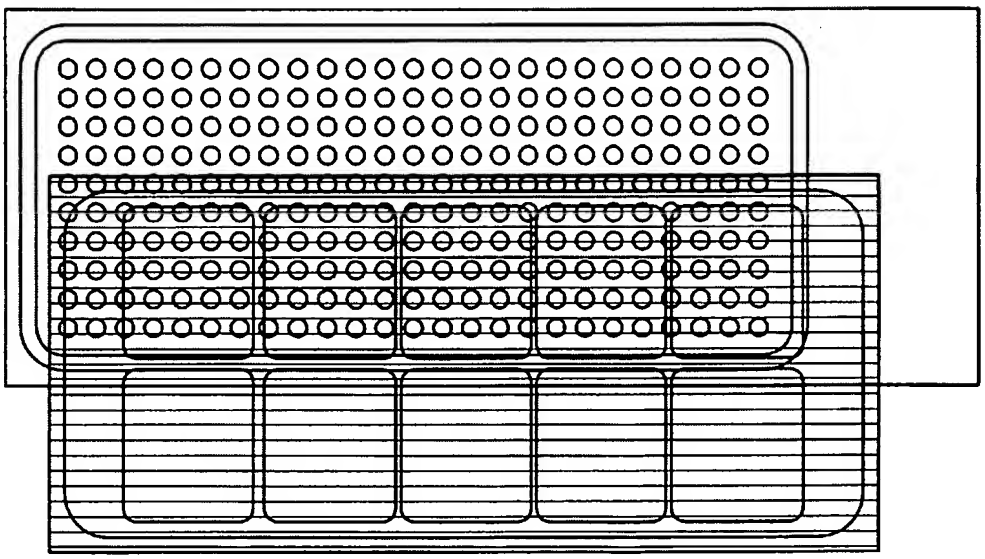
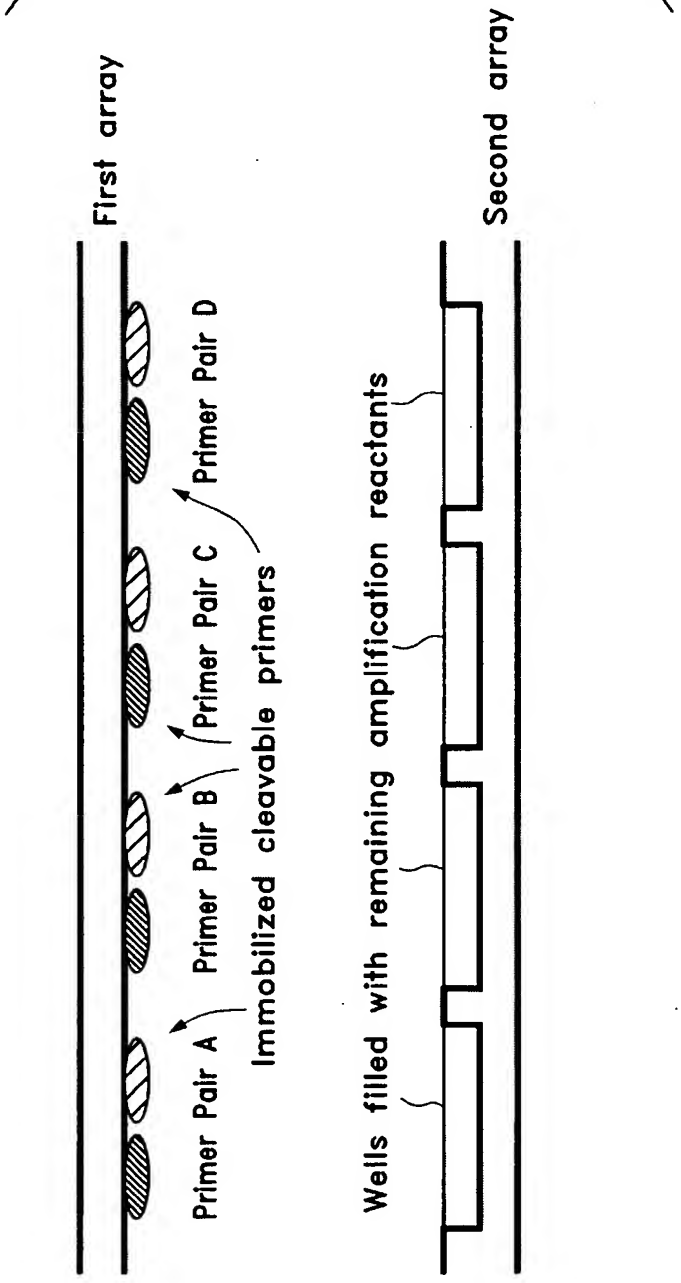


Fig. 3


First array
Second array
Primer Pair A
Primer Pair B
Primer Pair C
Primer Pair D
Immobilized cleavable primers
Wells filled with remaining amplification reactants

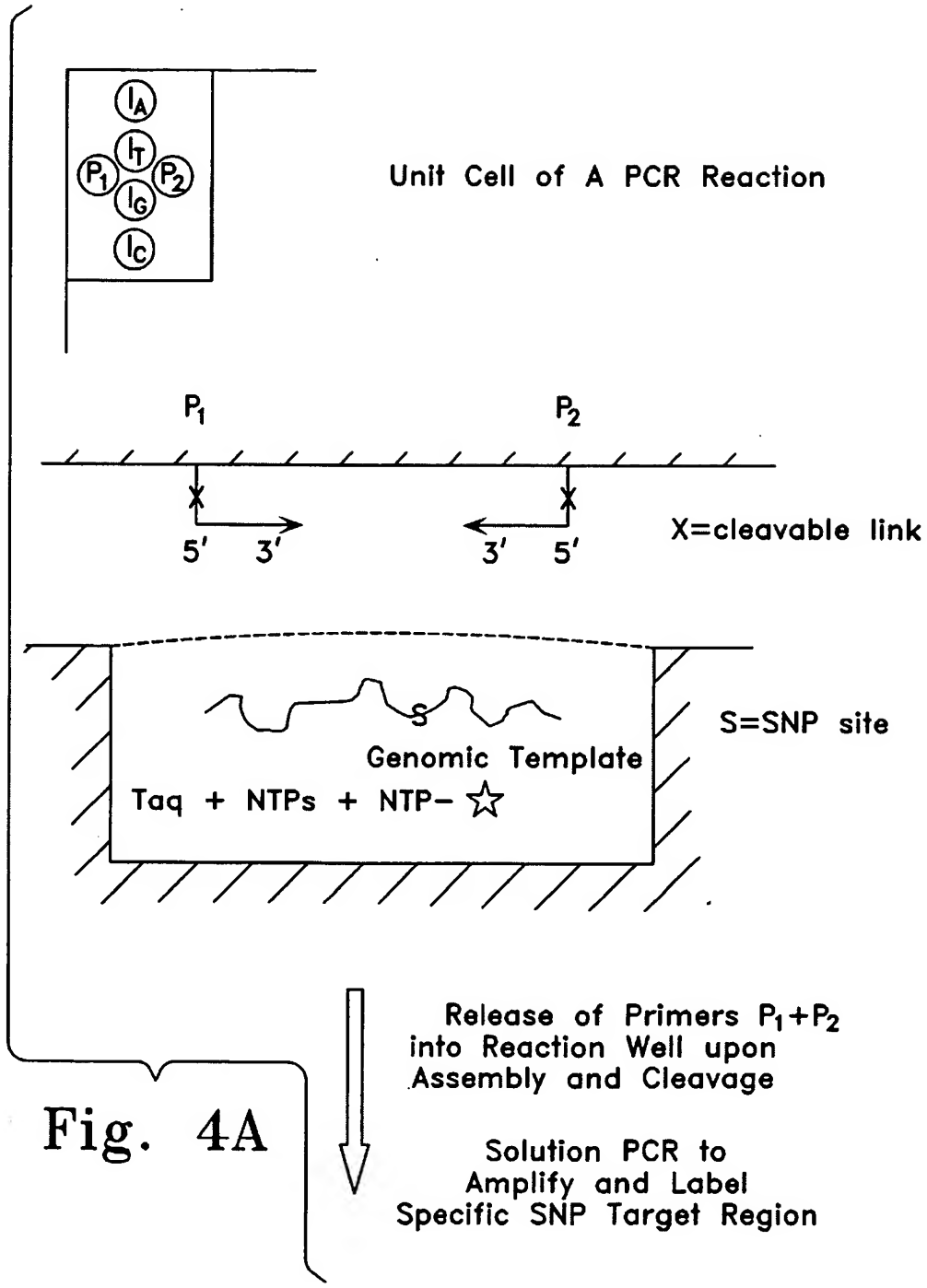

I_A
I_T
P_1
P_2
I_G
I_C
Unit Cell of A PCR Reaction
P_1
P_2
5′ 3′
3′ 5′
X=cleavable link
S
S=SNP site
Genomic Template
Taq + NTPs + NTP–
Release of Primers P_1+P_2 into Reaction Well upon Assembly and Cleavage
Solution PCR to Amplify and Label Specific SNP Target Region

Fig. 4A

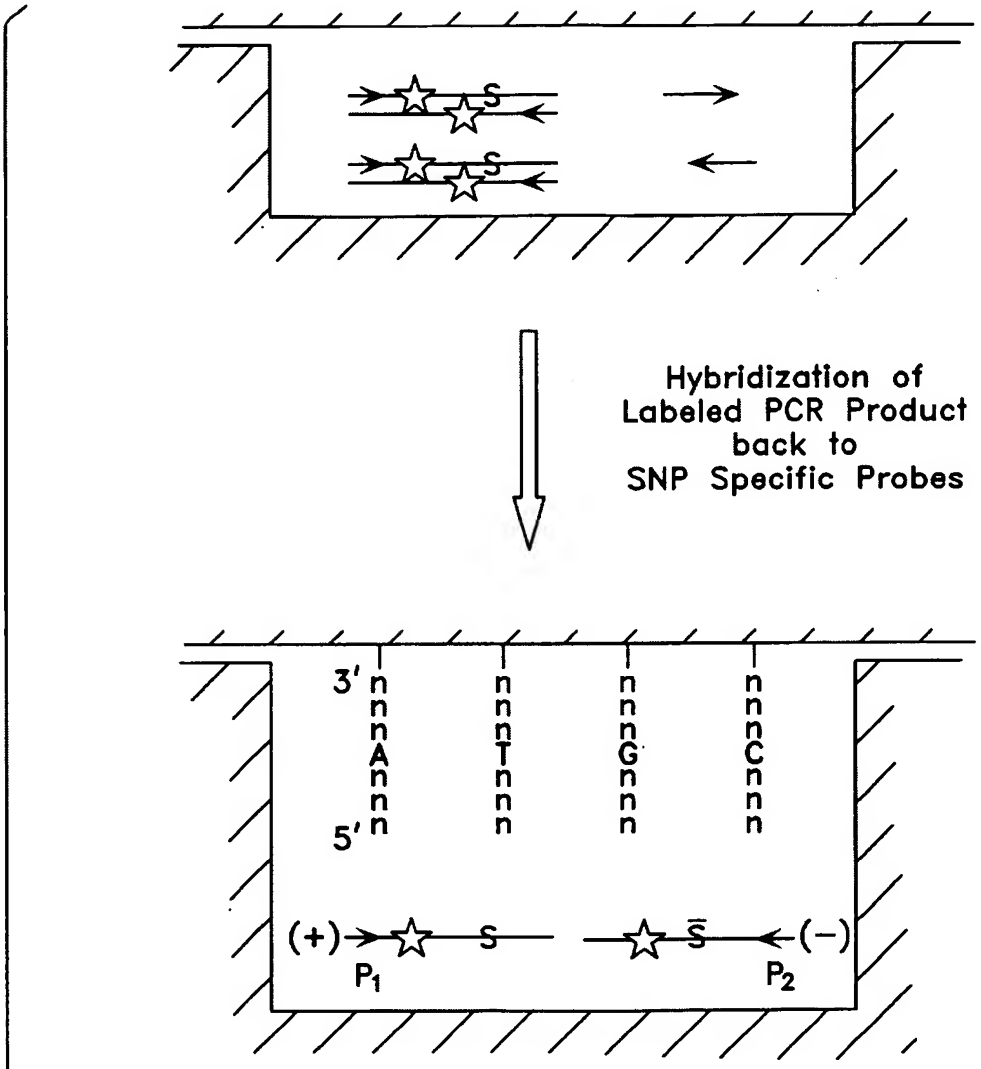


Fig. 4B

A/G pair interrogates (+) strand
T/C pair interrogates (−) strand

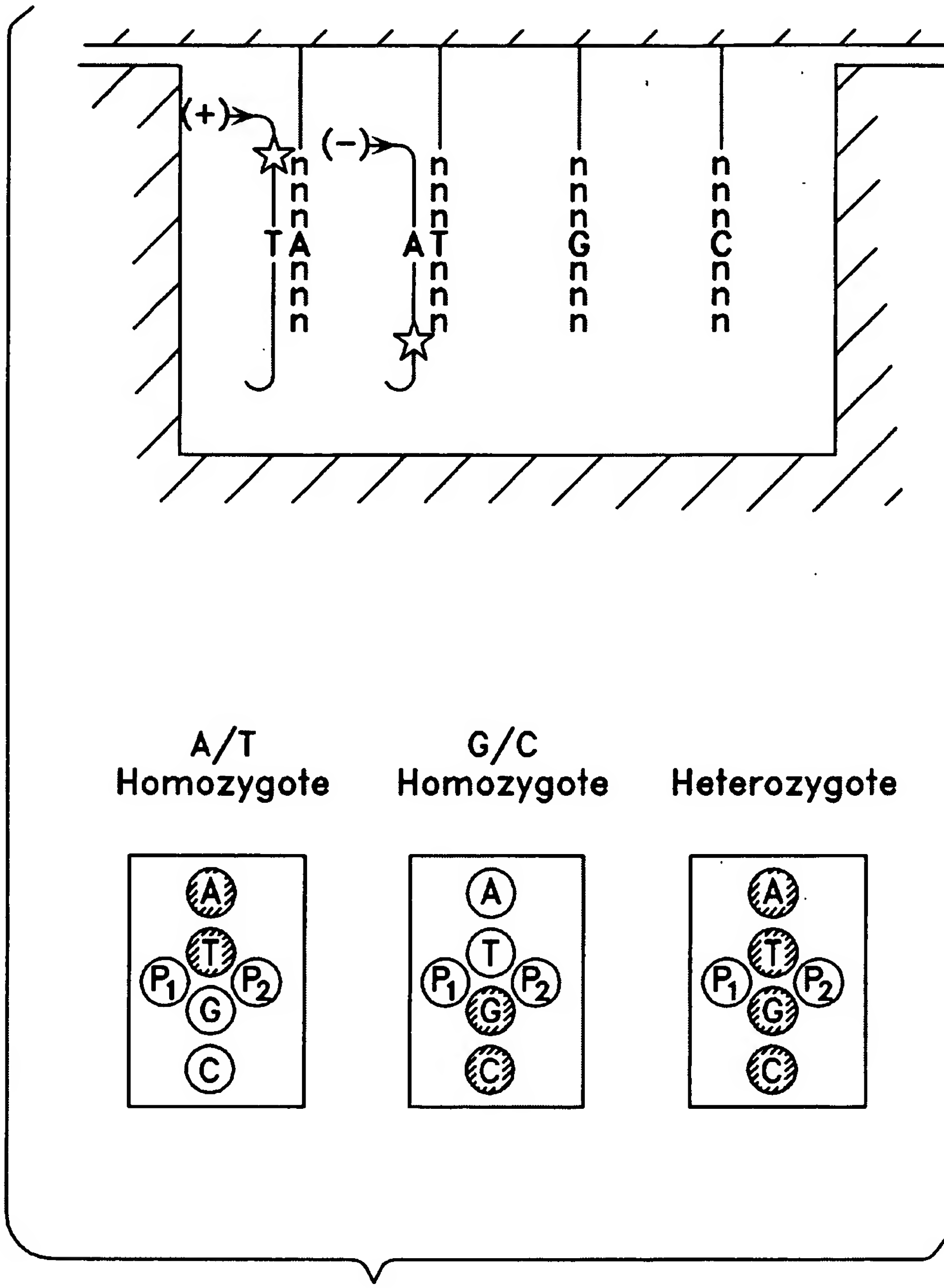

(+)
(−)
n
n
n
TA
n
n
n
n
n
n
AT
n
n
n
n
n
n
G
n
n
n
n
n
n
C
n
n
n
A/T
Homozygote
G/C
Homozygote
Heterozygote
A
T
P1
G
P2
C

Fig. 4C

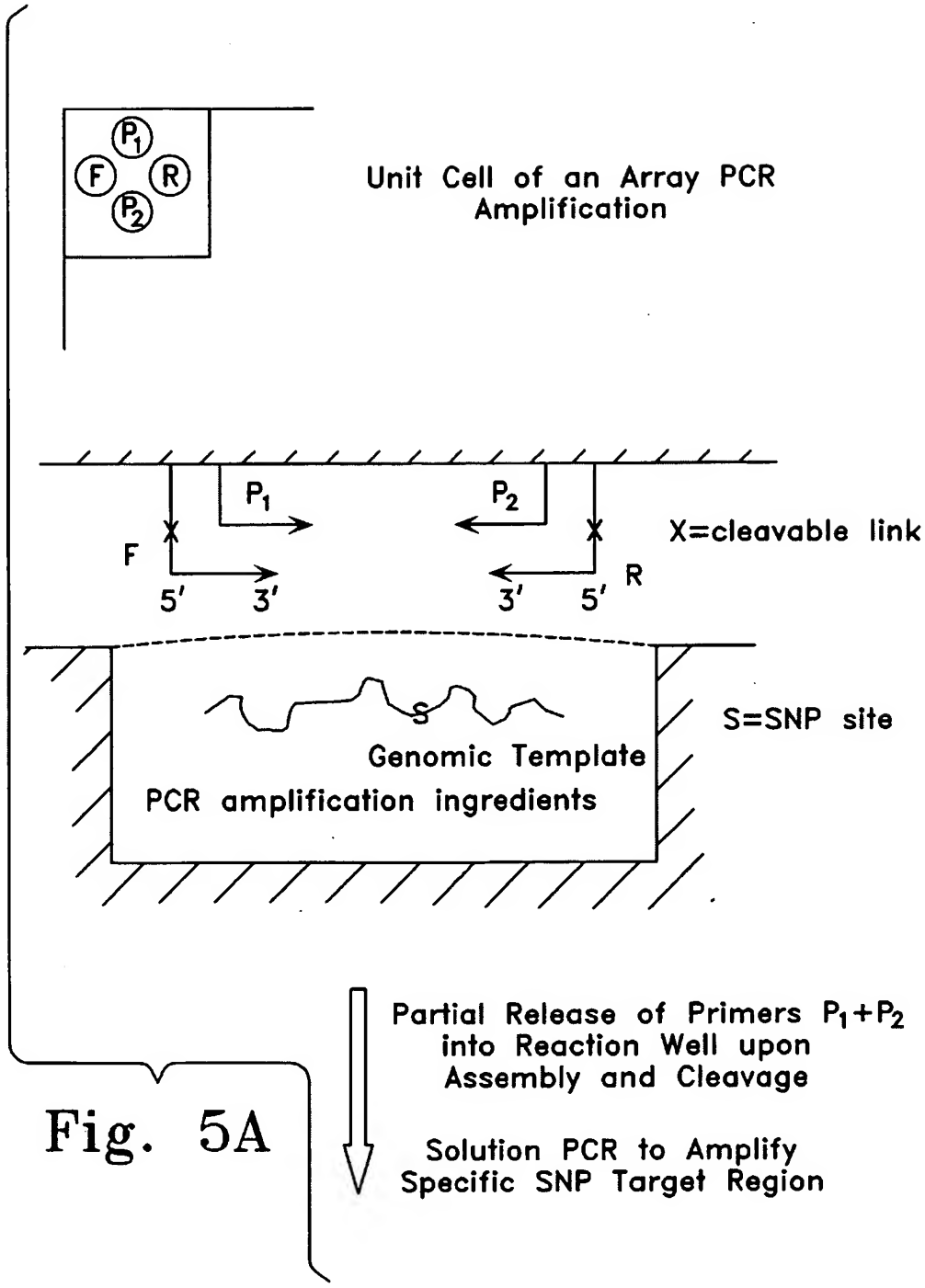

P1
F
R
P2
Unit Cell of an Array PCR Amplification
P1
P2
X=cleavable link
F
R
5'
3'
3'
5'
S
S=SNP site
Genomic Template
PCR amplification ingredients
Partial Release of Primers P1+P2 into Reaction Well upon Assembly and Cleavage
Solution PCR to Amplify Specific SNP Target Region

Fig. 5A

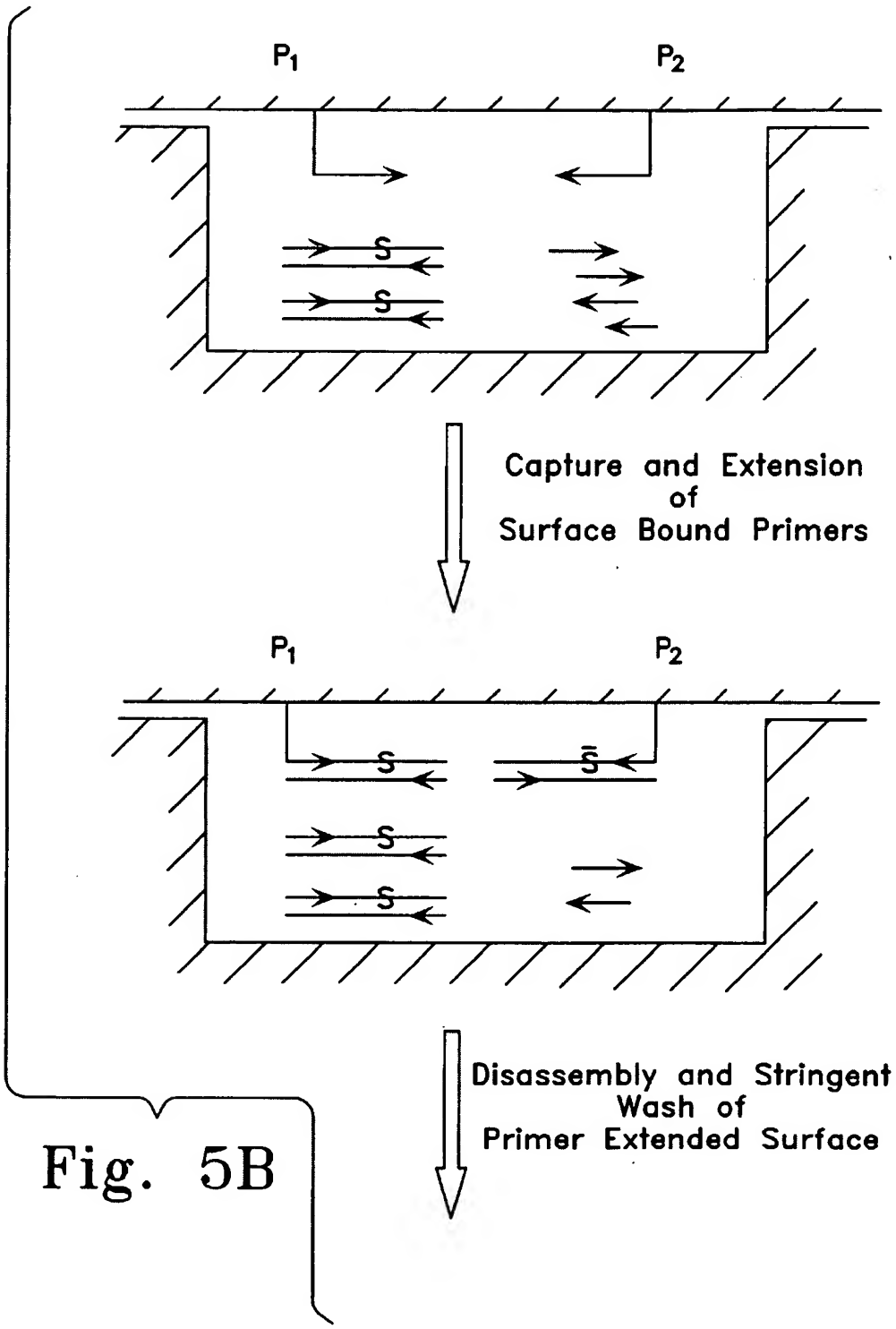

P1
P2
S
S
Capture and Extension
of
Surface Bound Primers
P1
P2
S
S̄
S
S
Disassembly and Stringent
Wash of
Primer Extended Surface

Fig. 5B

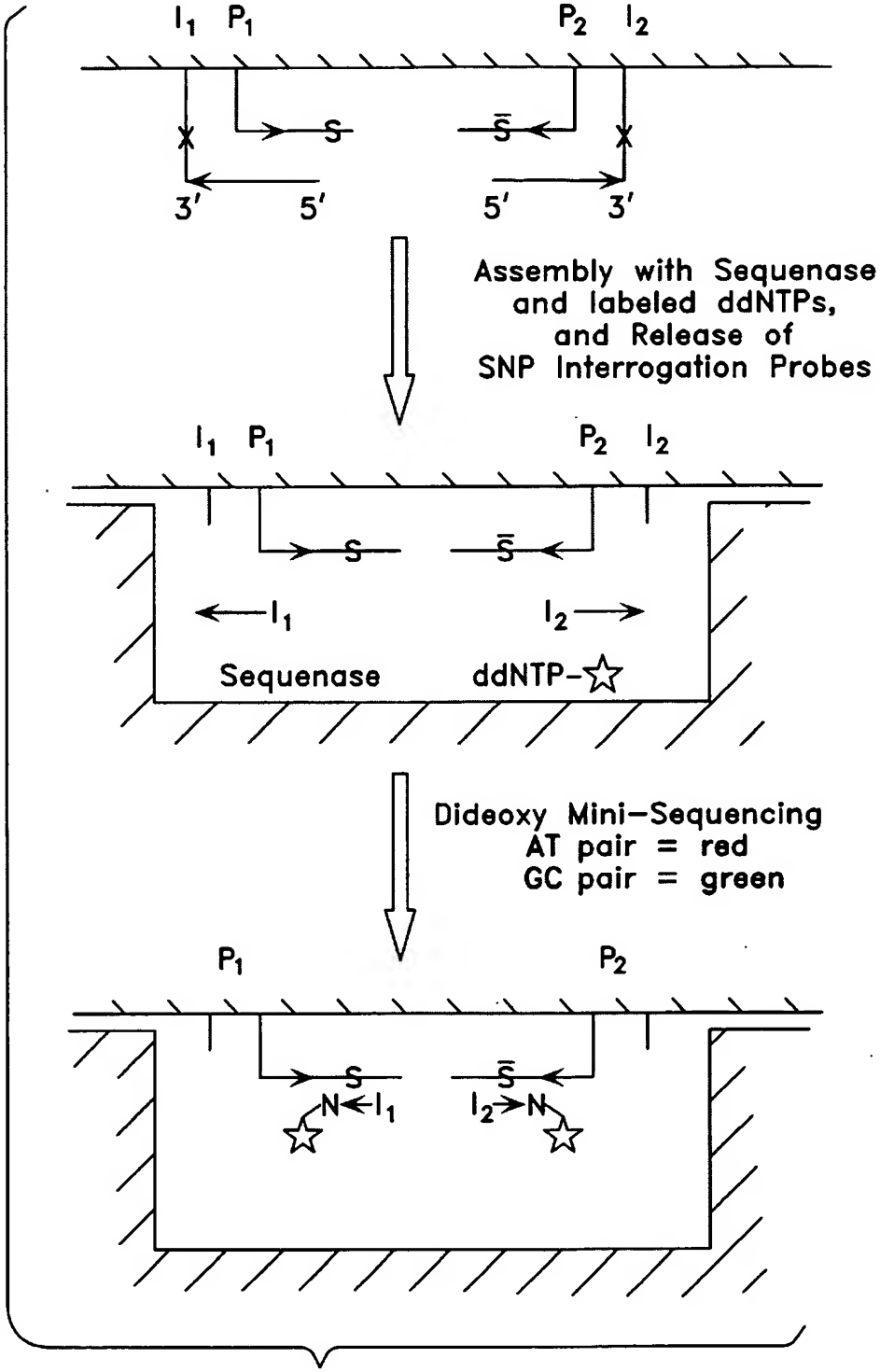

I1 P1
P2 I2
S
S̄
3′
5′
5′
3′
Assembly with Sequenase
and labeled ddNTPs,
and Release of
SNP Interrogation Probes
I1 P1
P2 I2
S
S̄
I1
I2
Sequenase
ddNTP-☆
Dideoxy Mini-Sequencing
AT pair = red
GC pair = green
P1
P2
S
S̄
N←I1
I2→N

Fig. 5C

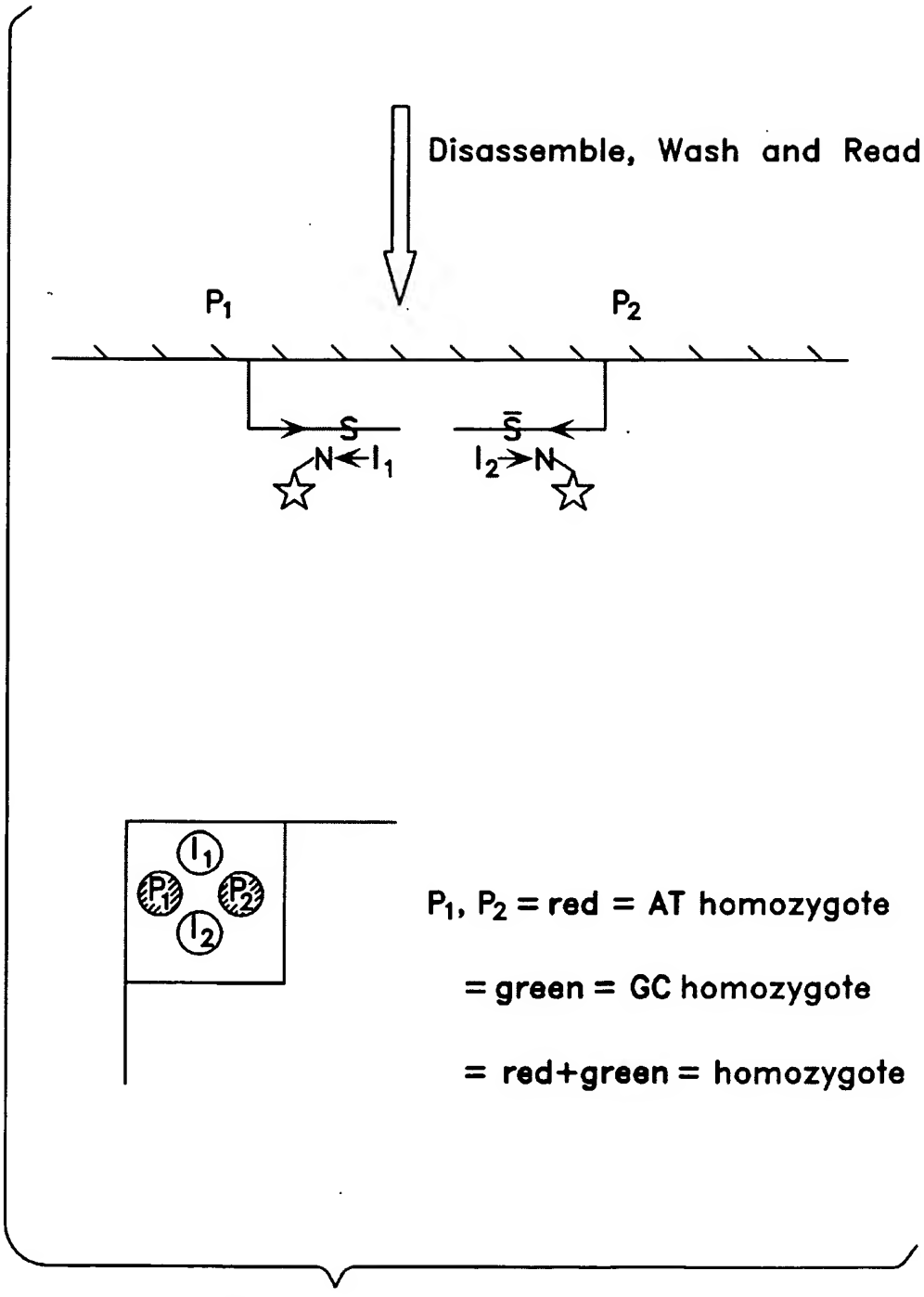

Disassemble, Wash and Read
P1
P2
S
N←I1
S̄
I2→N
I1
P1
P2
I2
P1, P2 = red = AT homozygote
= green = GC homozygote
= red+green = homozygote

Fig. 5D

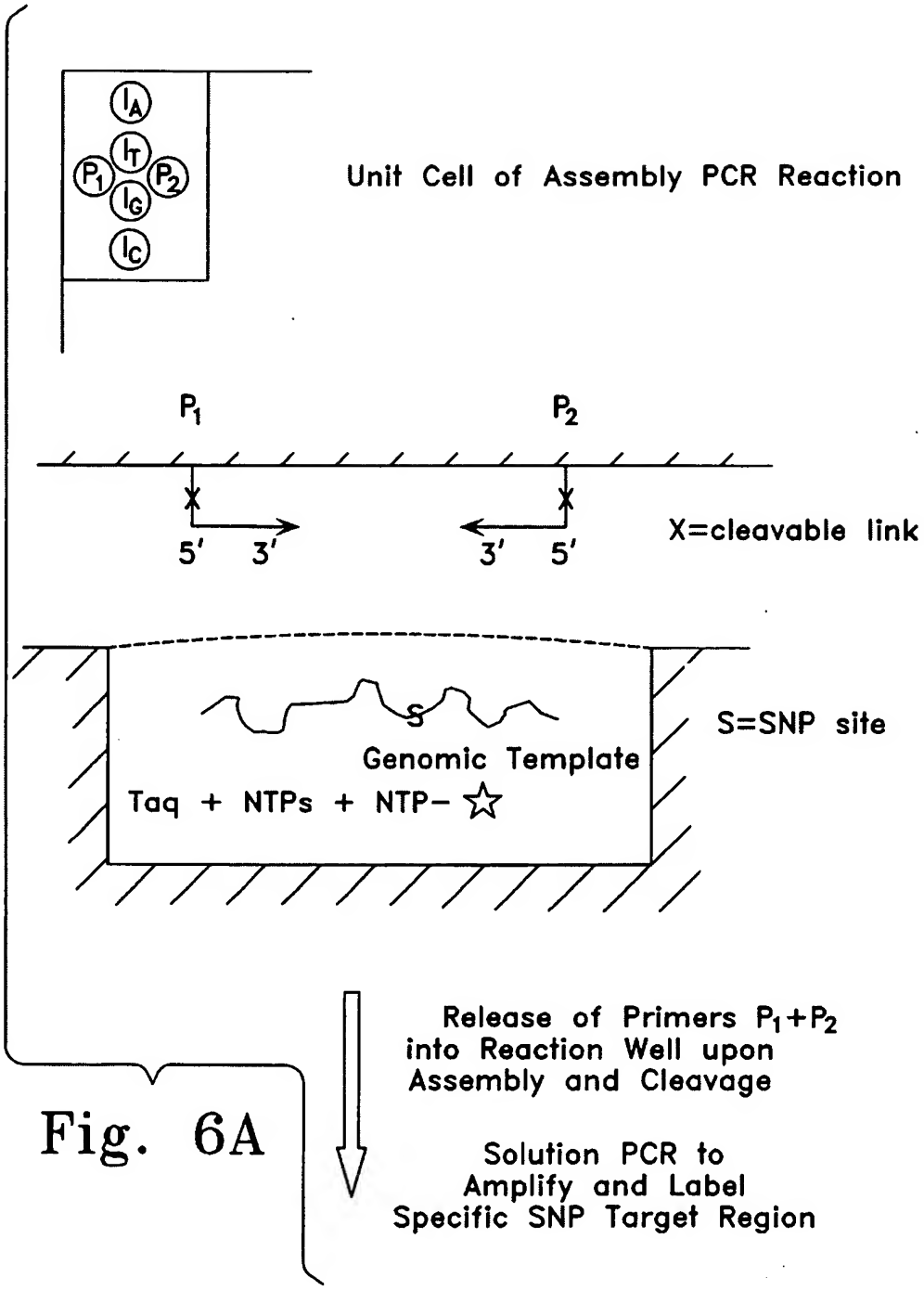

I_A
I_T
P_1
P_2
I_G
I_C
Unit Cell of Assembly PCR Reaction
P_1
P_2
5′ 3′
3′ 5′
X=cleavable link
S
S=SNP site
Genomic Template
Taq + NTPs + NTP–
Release of Primers P_1+P_2 into Reaction Well upon Assembly and Cleavage
Solution PCR to Amplify and Label Specific SNP Target Region

Fig. 6A

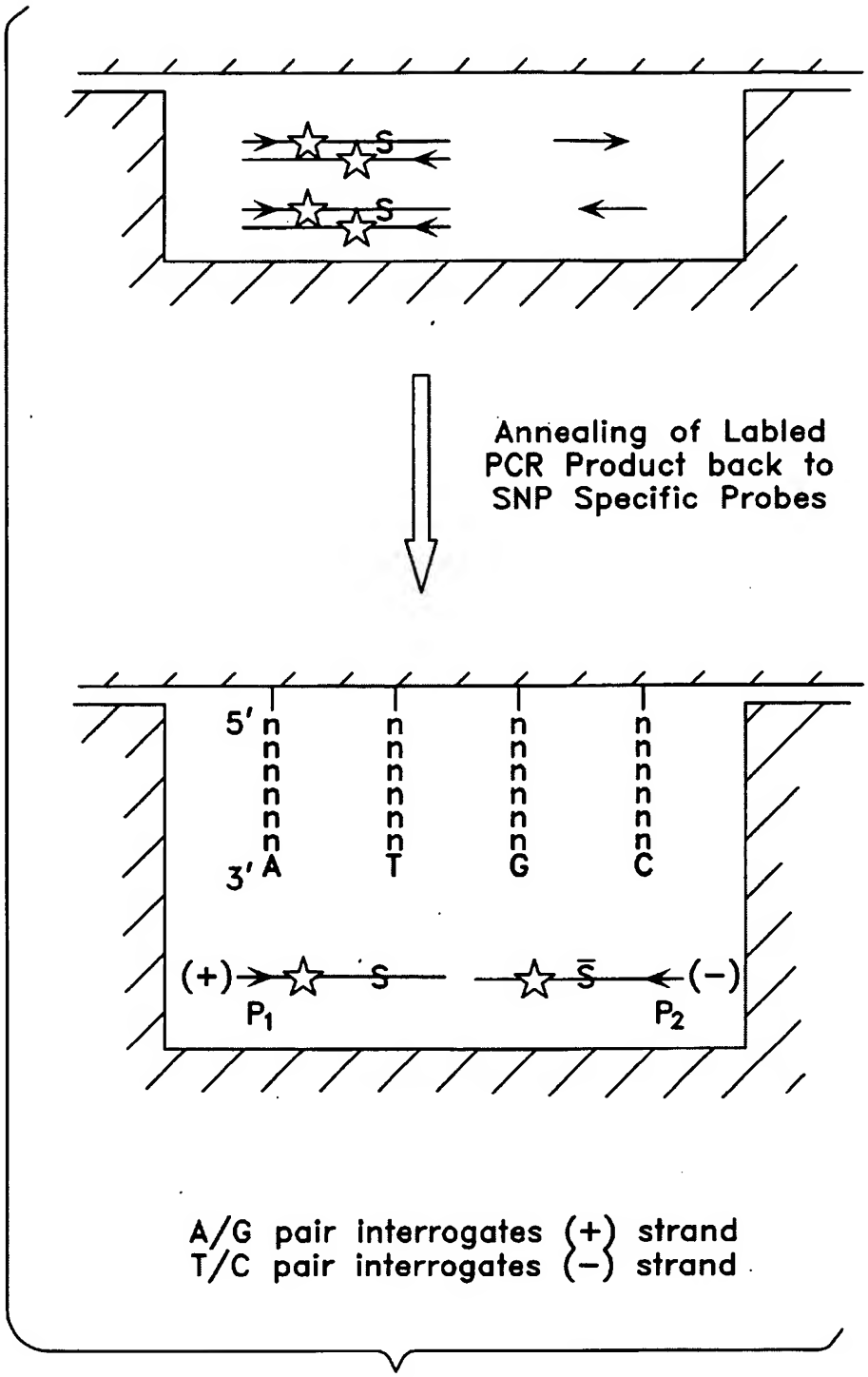

S
S
Annealing of Labled
PCR Product back to
SNP Specific Probes
5'
3'
A
T
G
C
(+)
S
P1
S
(−)
P2
A/G pair interrogates (+) strand
T/C pair interrogates (−) strand

Fig. 6B

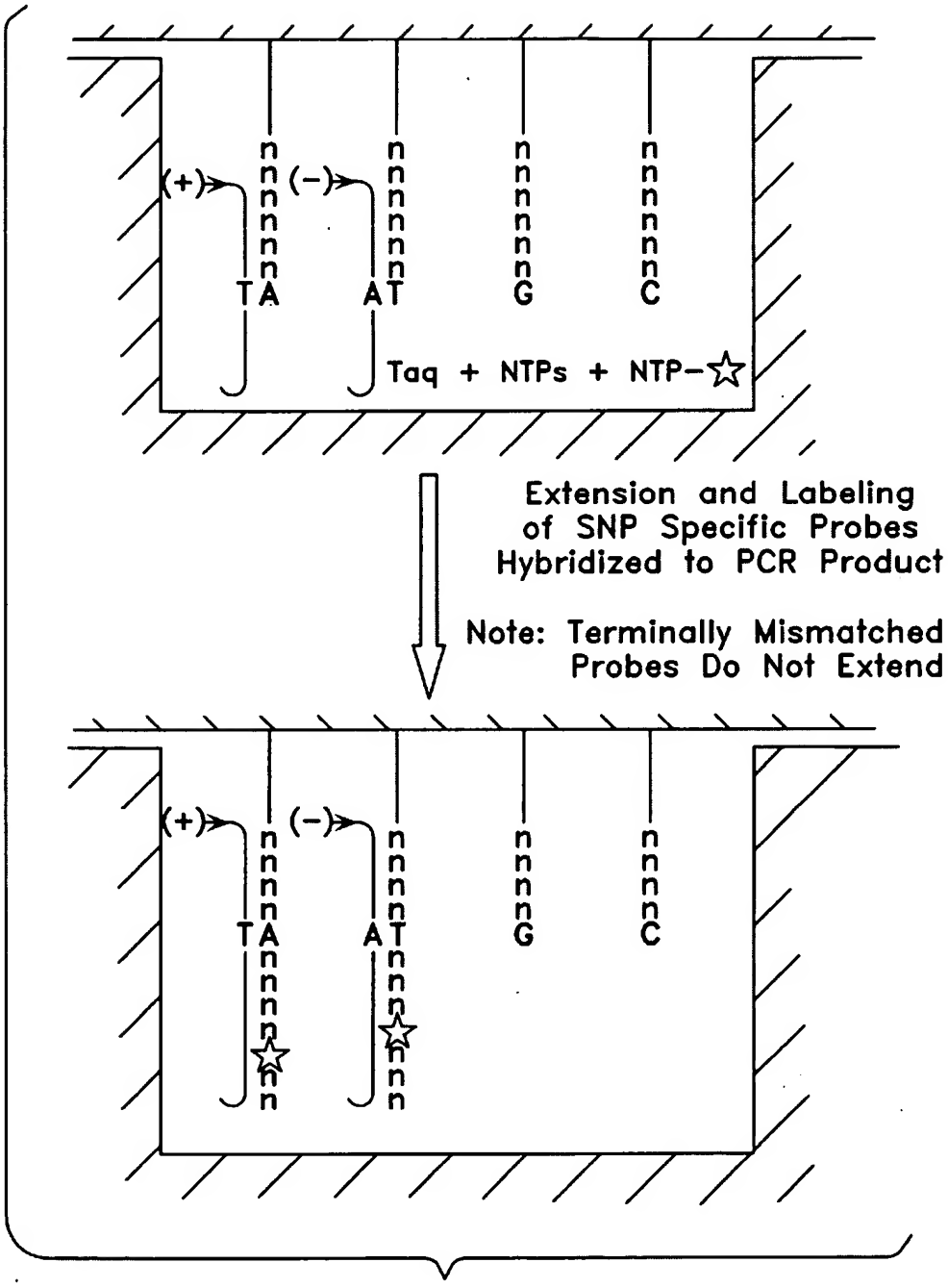

(+)
(-)
TA
AT
G
C
Taq + NTPs + NTP-☆
Extension and Labeling
of SNP Specific Probes
Hybridized to PCR Product
Note: Terminally Mismatched
Probes Do Not Extend

Fig. 6C

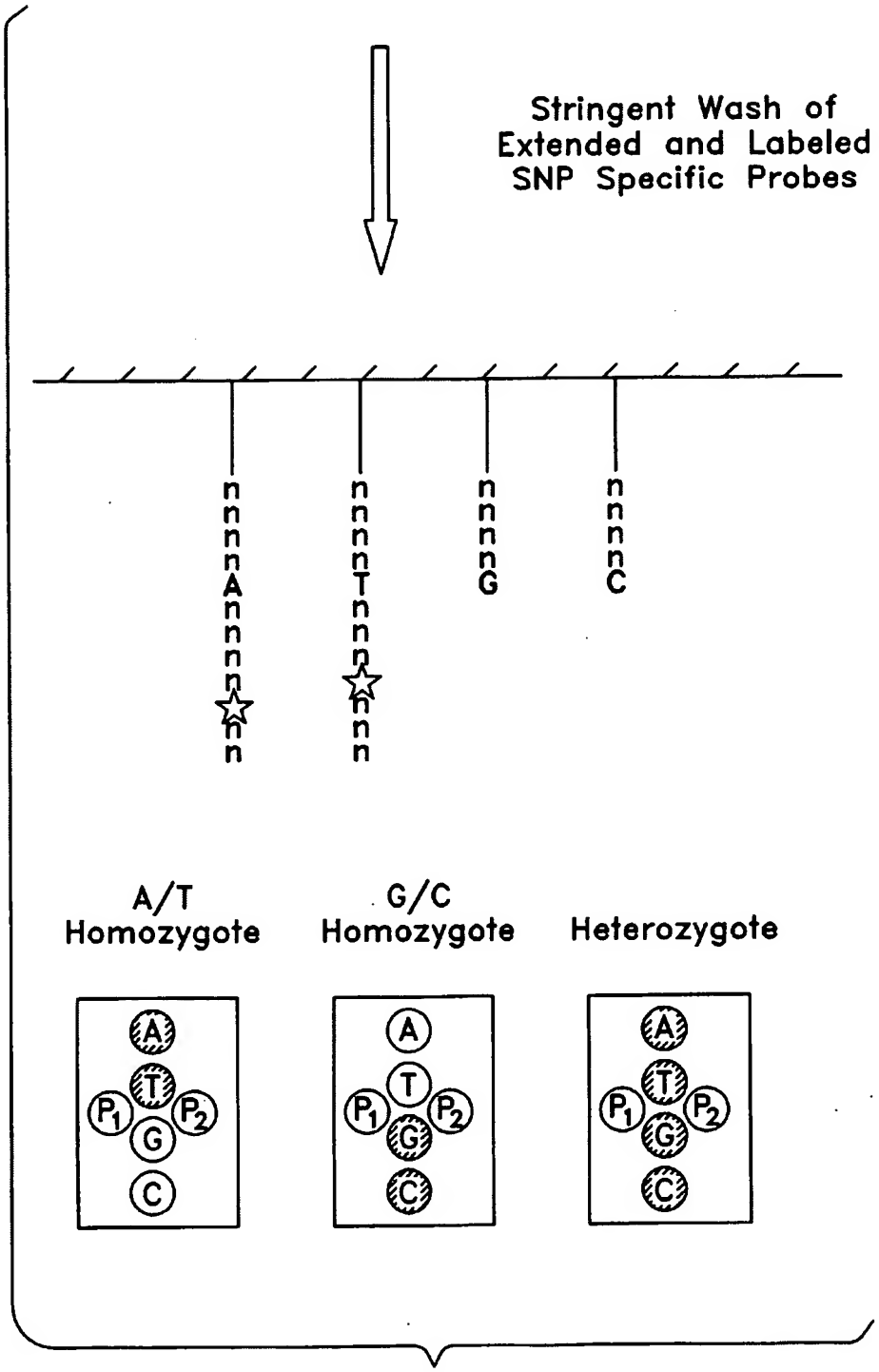

Stringent Wash of
Extended and Labeled
SNP Specific Probes
nnnnAnnnnnn
nnnnTnnnnnn
nnnnG
nnnnC
A/T
Homozygote
G/C
Homozygote
Heterozygote
A
T
P1
P2
G
C

Fig. 6D

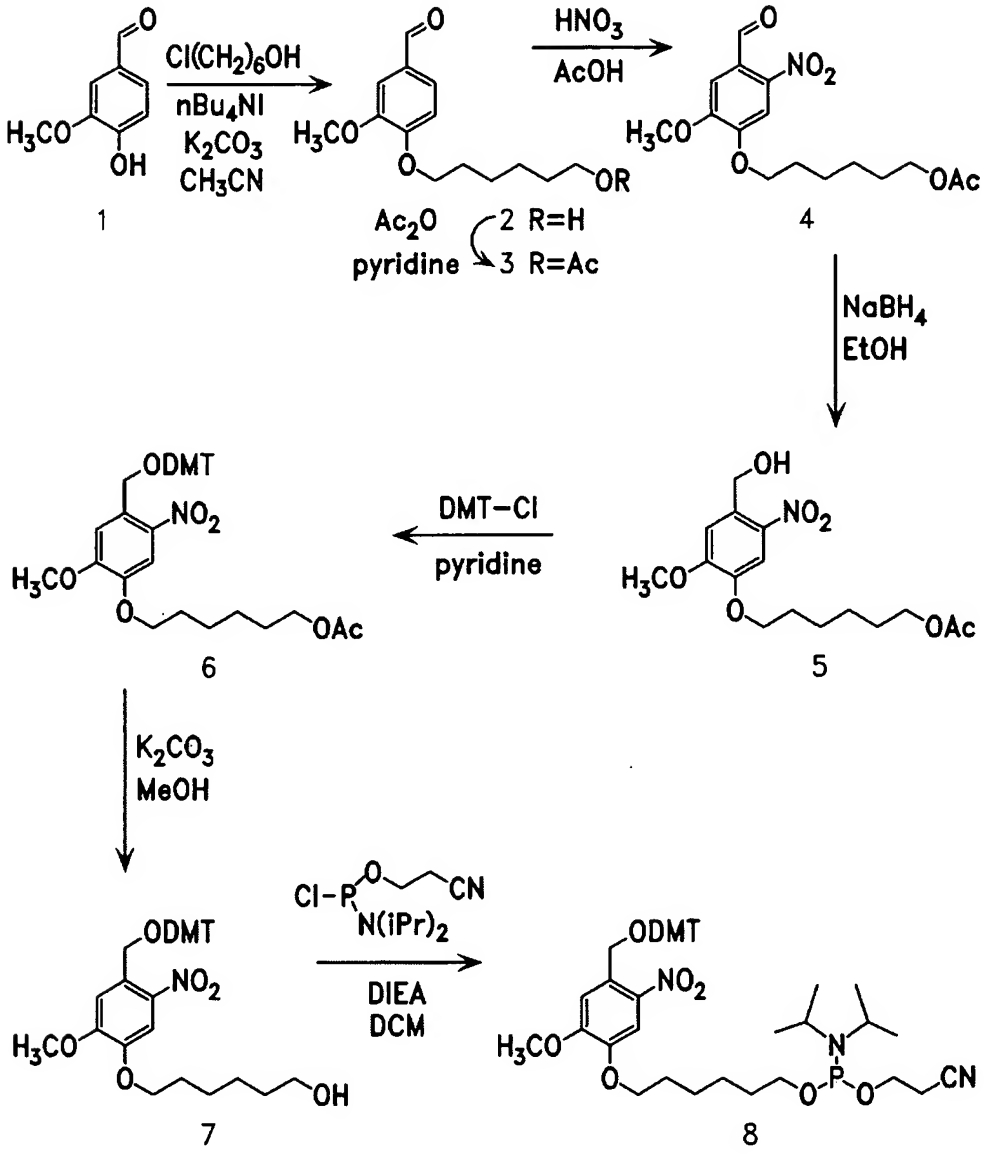

Cl(CH2)6OH
nBu4NI
K2CO3
CH3CN
HNO3
AcOH
Ac2O
pyridine
2 R=H
3 R=Ac
NaBH4
EtOH
DMT-Cl
pyridine
K2CO3
MeOH
N(iPr)2
DIEA
DCM
ODMT
NO2
OAc
OH
CN

Fig. 7

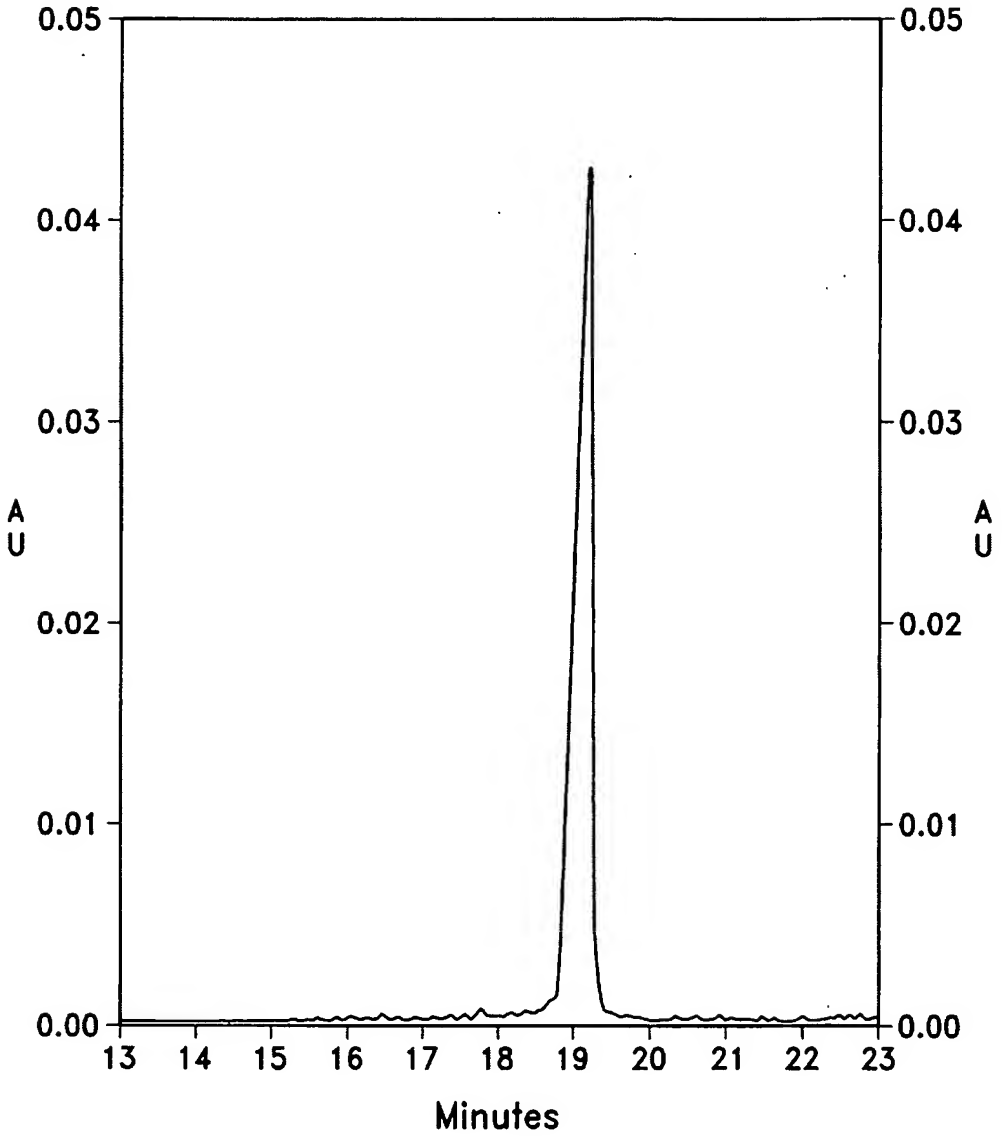


Fig. 8

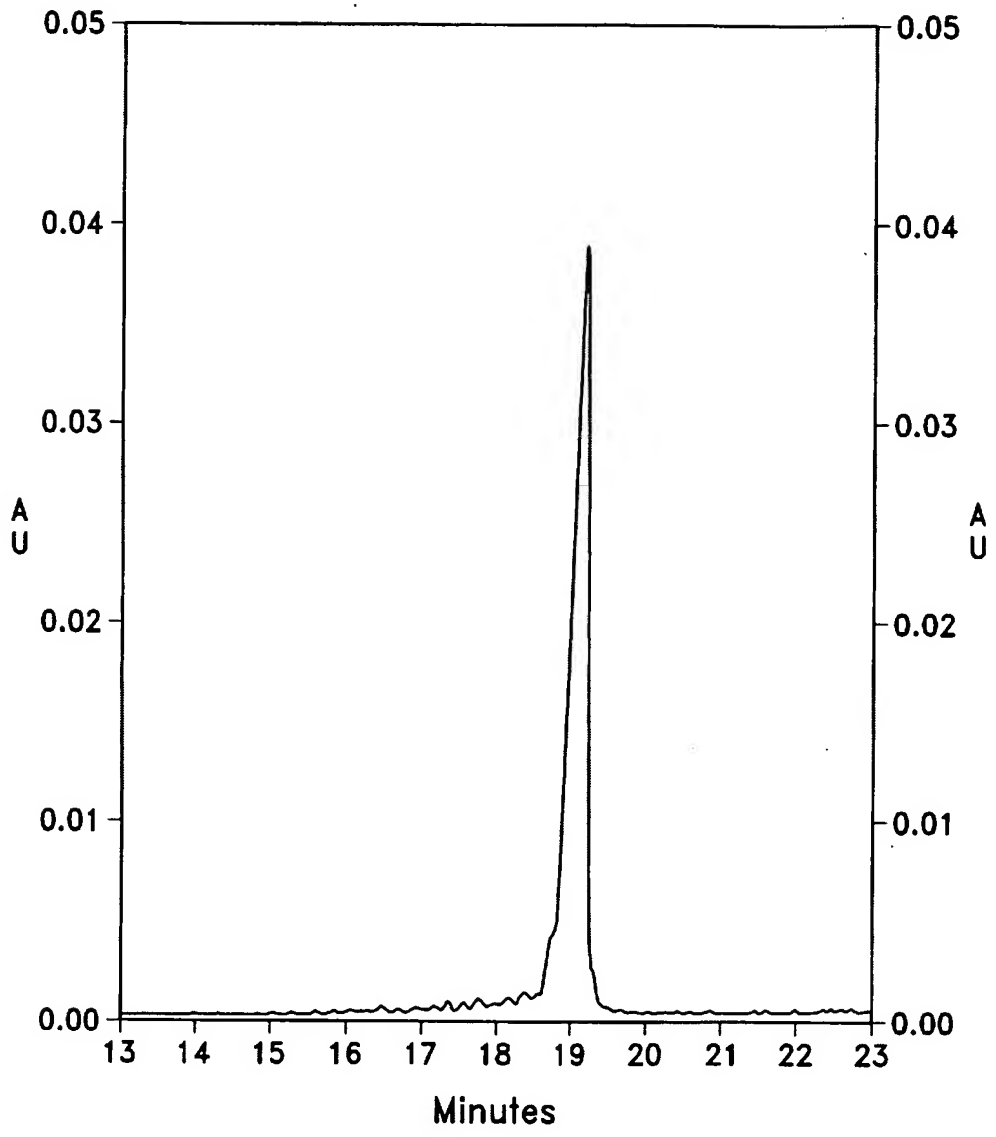


Fig. 9

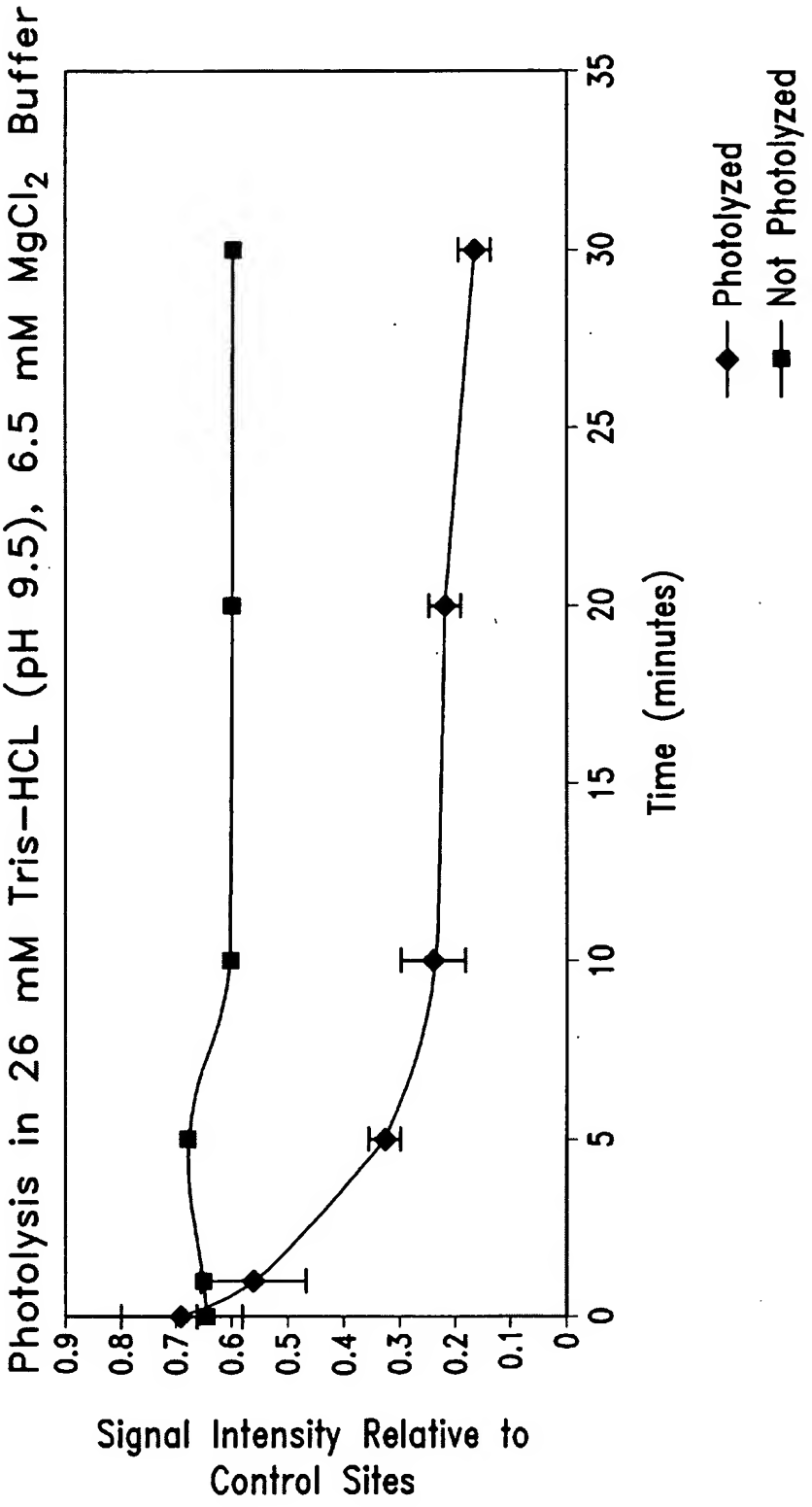

Photolysis in 26 mM Tris–HCL (pH 9.5), 6.5 mM $MgCl_2$ Buffer
Signal Intensity Relative to Control Sites
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
5
10
15
20
25
30
35
Time (minutes)
Photolyzed
Not Photolyzed

Fig. 10

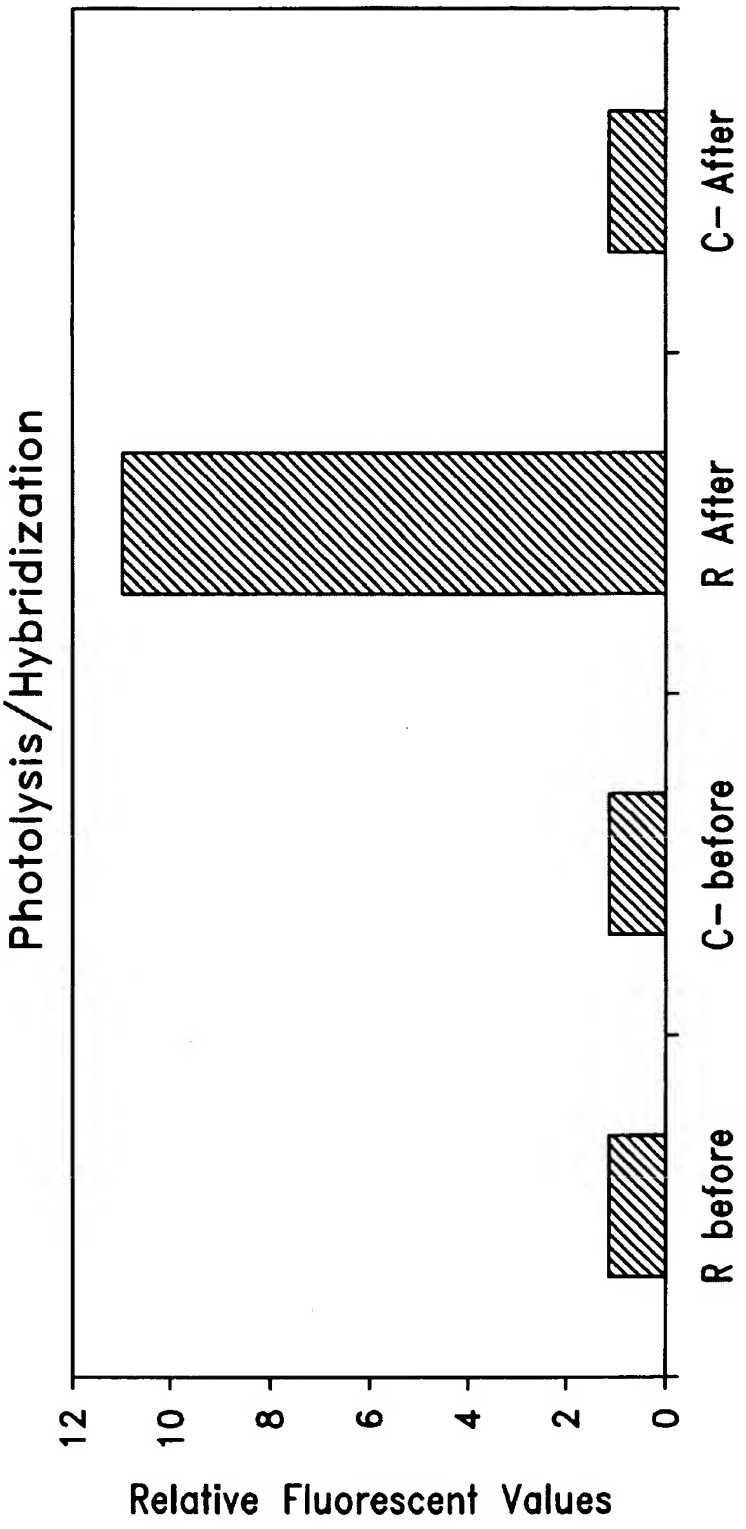

Photolysis/Hybridization
12
10
8
6
4
2
0
Relative Fluorescent Values
R before
C- before
R After
C- After
Feature Identification

Fig. 11

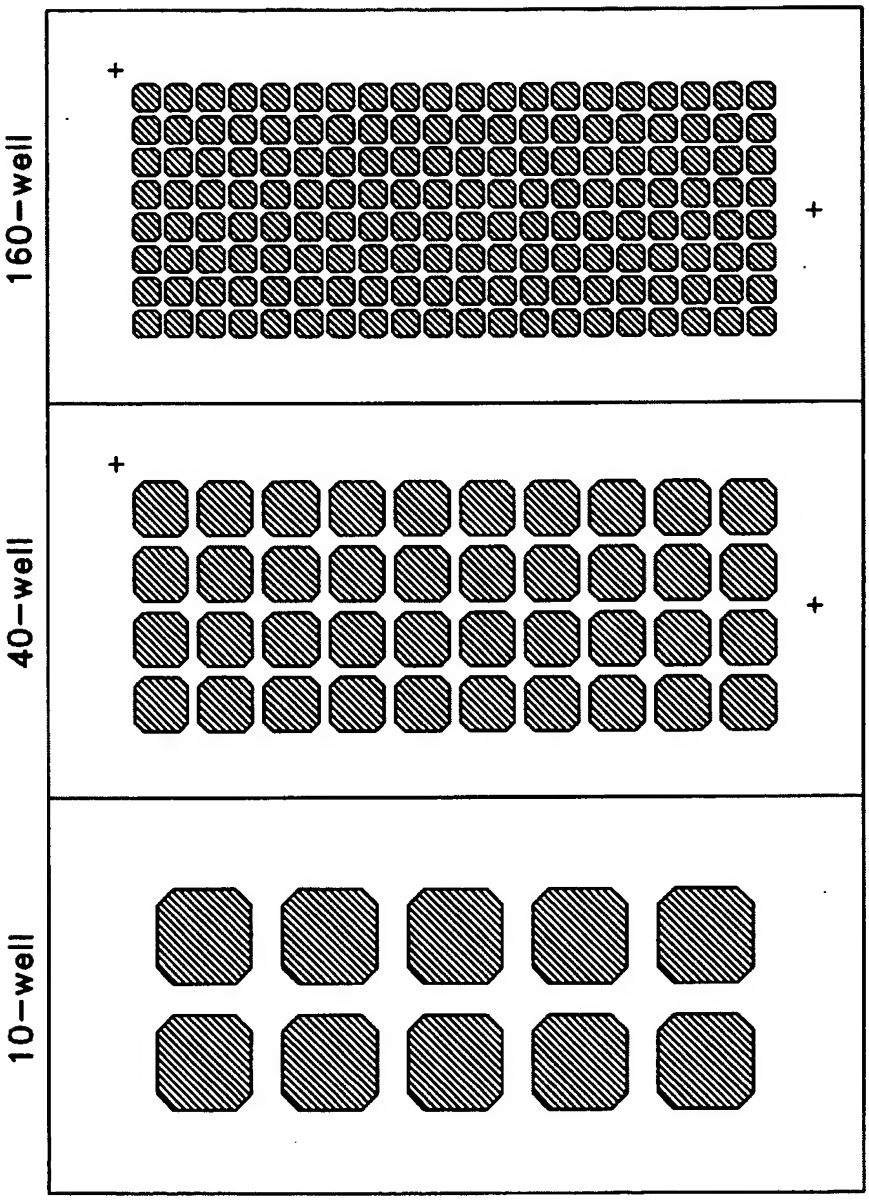
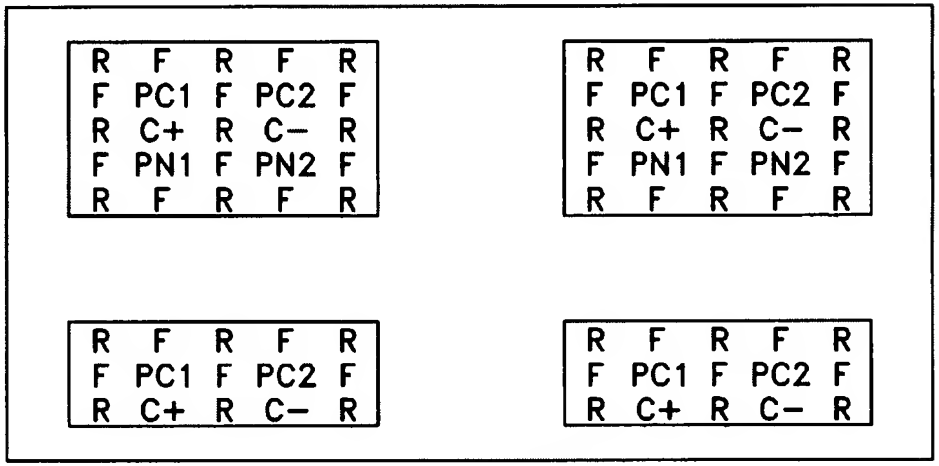


Fig. 12



Probe	Description	ID	Name	3–5 Sequence
F	Forward primer	FPC	F191RevCmp	ctcaacccgaatctccgataaaa
R	Reverse primer	RPRC	R191RevCmp	cgggtcatgtcttcaactaactg
PC1	Probe coding 1	A	191C_A	CTTTGGACCCACCCA
PC2	Probe coding 2	G	191C_G	CTTTGGGCCCACC
PN1	Probe non-coding 1	T	191NC_T	TGGGTGGGTCCAAAGAA
PN2	Probe non-coding 2	C	191NC_C	TGGGTGGGCCCAAAGAA
C+	Positive Control	+	F191Rev	TTTTATCGGAGATTCGGGTTGAG
C−	Negative Control	−	F481Rev	GGAGAGAGAAGACAGTTCGTCTT

Fig. 13

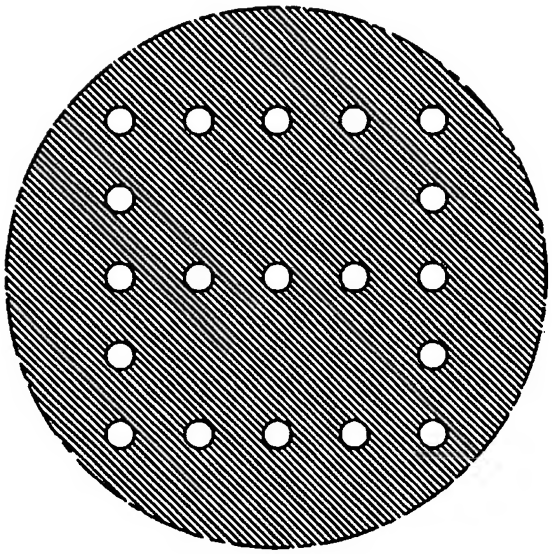


Fig. 14A

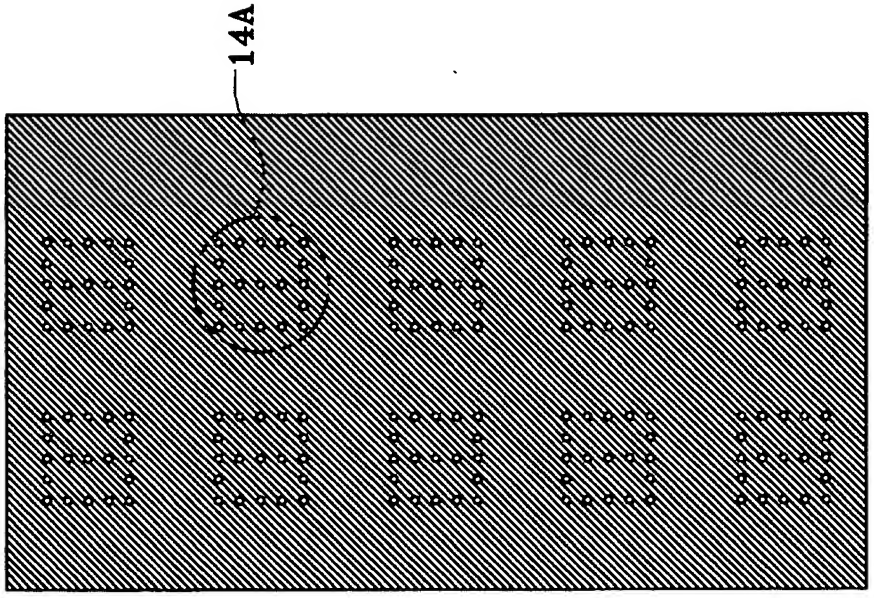


Fig. 14

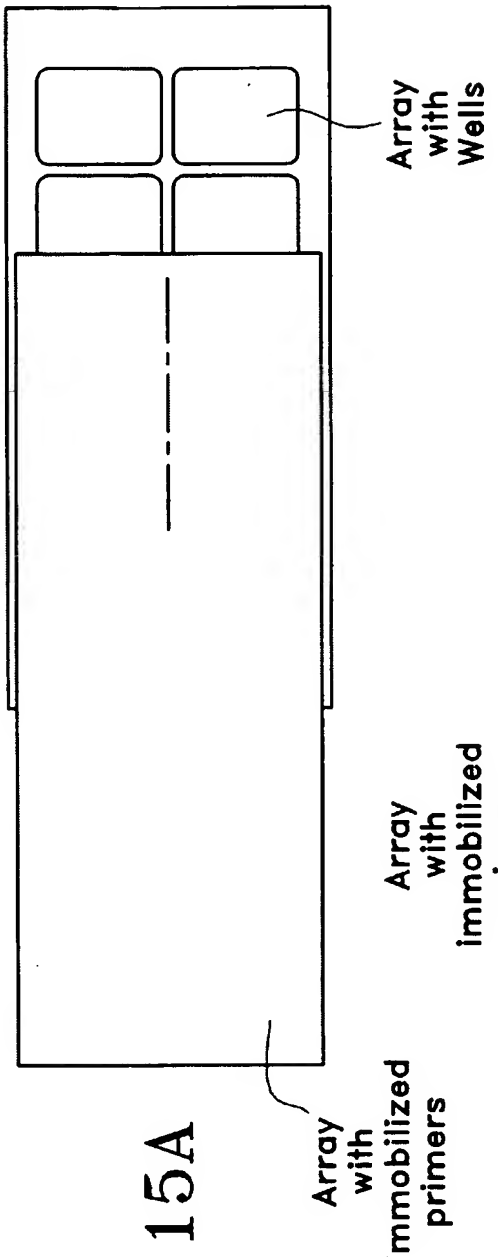


Fig. 15A

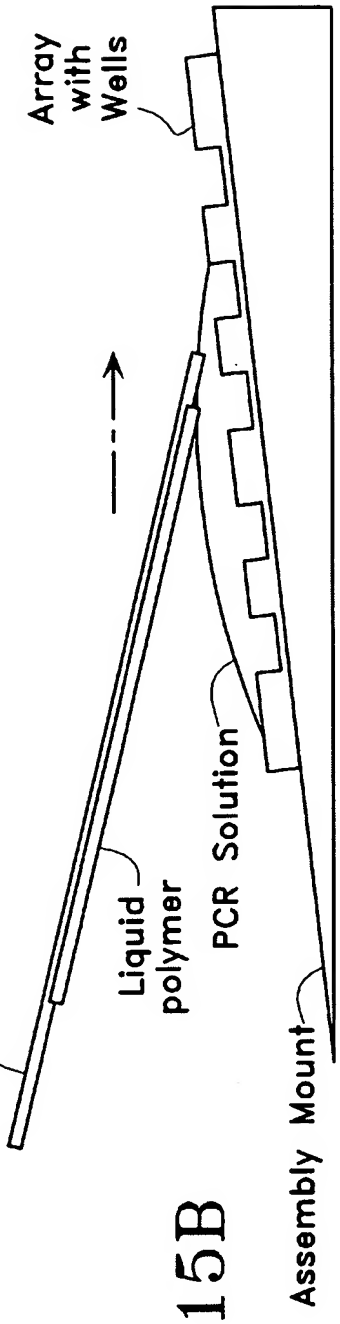


Fig. 15B

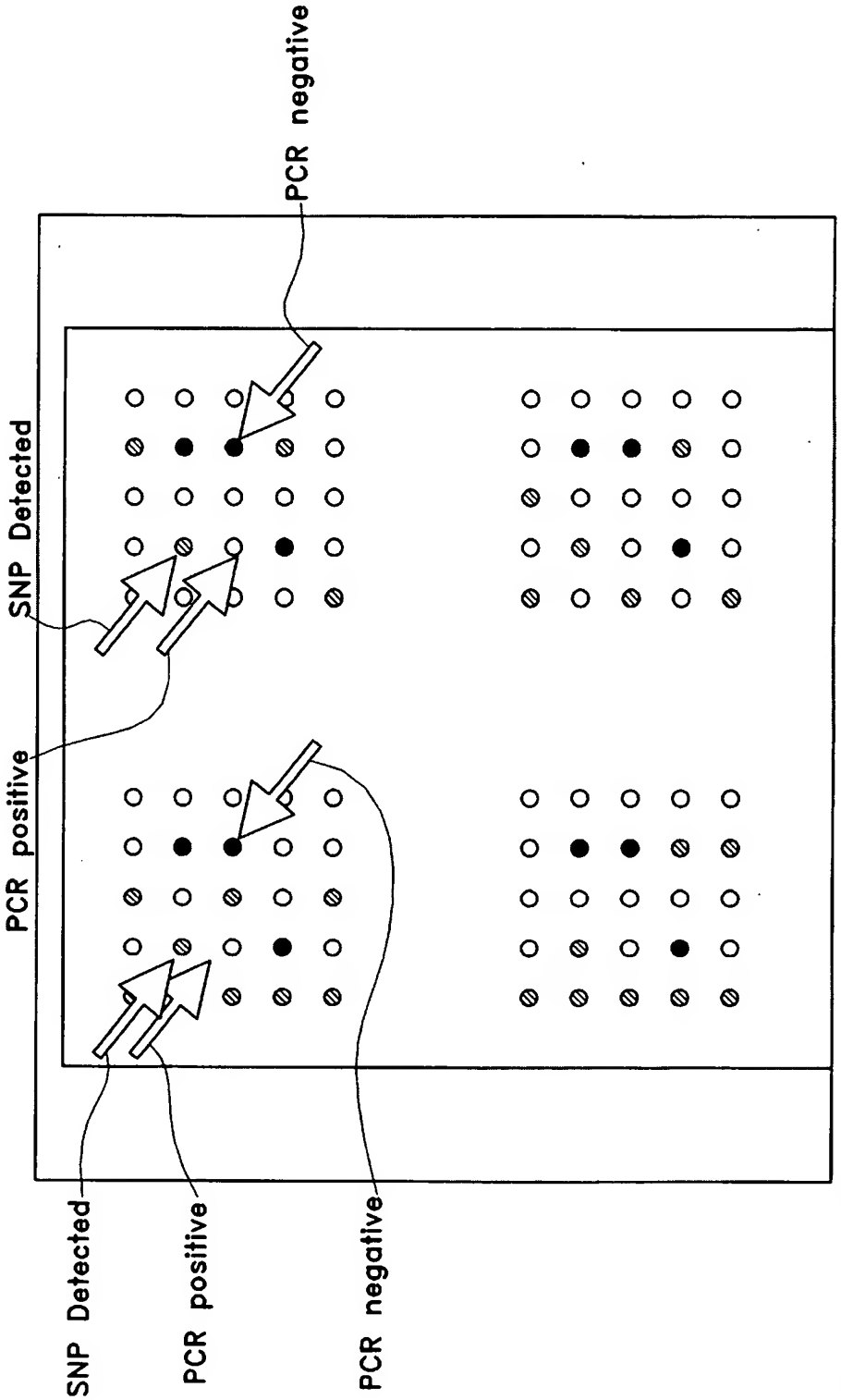

PCR negative
SNP Detected
PCR positive
SNP Detected
PCR positive
PCR negative

Fig. 16

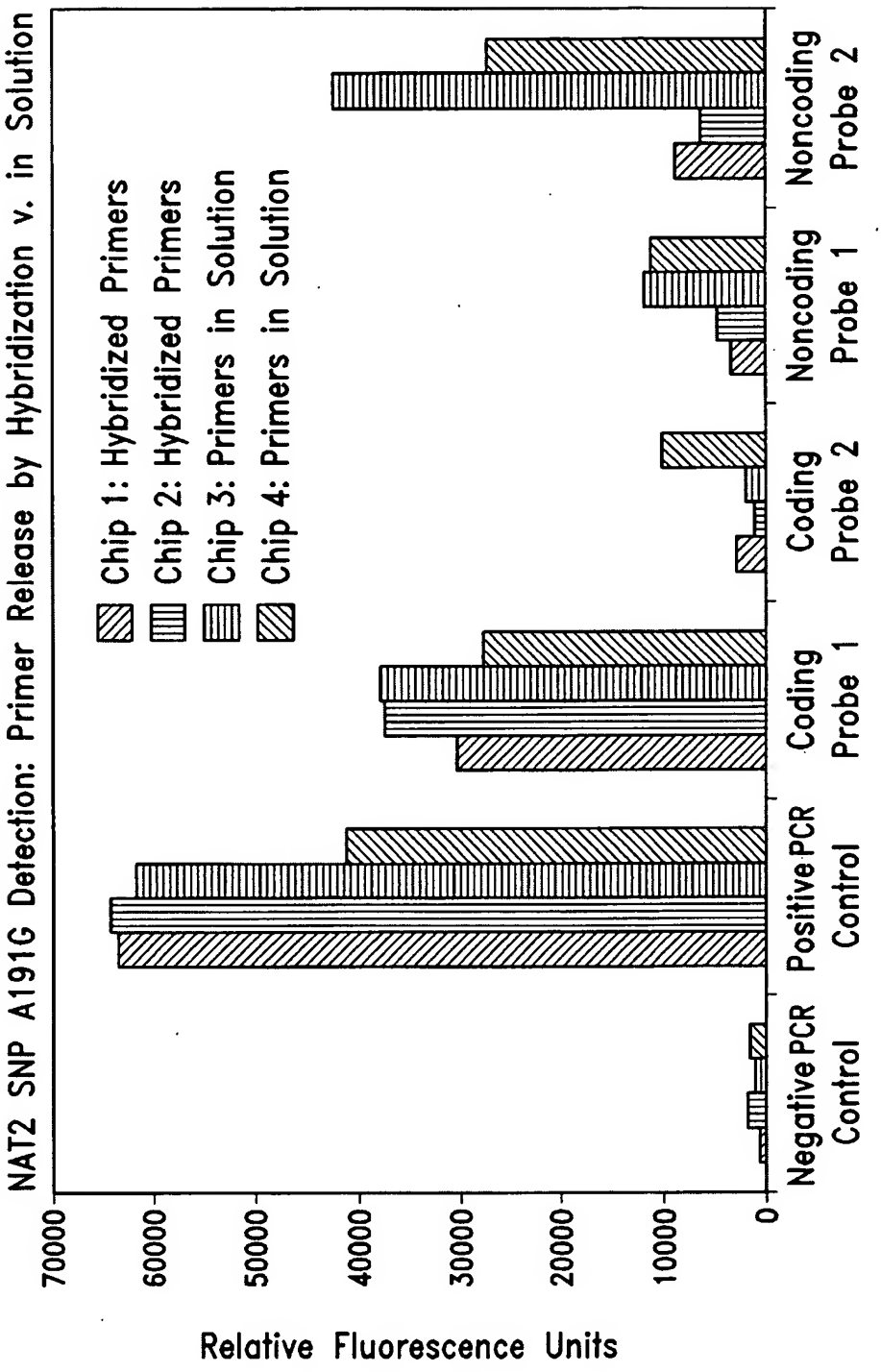

NAT2 SNP A191G Detection: Primer Release by Hybridization v. in Solution
Chip 1: Hybridized Primers
Chip 2: Hybridized Primers
Chip 3: Primers in Solution
Chip 4: Primers in Solution
70000
60000
50000
40000
30000
20000
10000
0
Relative Fluorescence Units
Negative PCR Control
Positive PCR Control
Coding Probe 1
Coding Probe 2
Noncoding Probe 1
Noncoding Probe 2
Flexchip Feature

Fig. 17

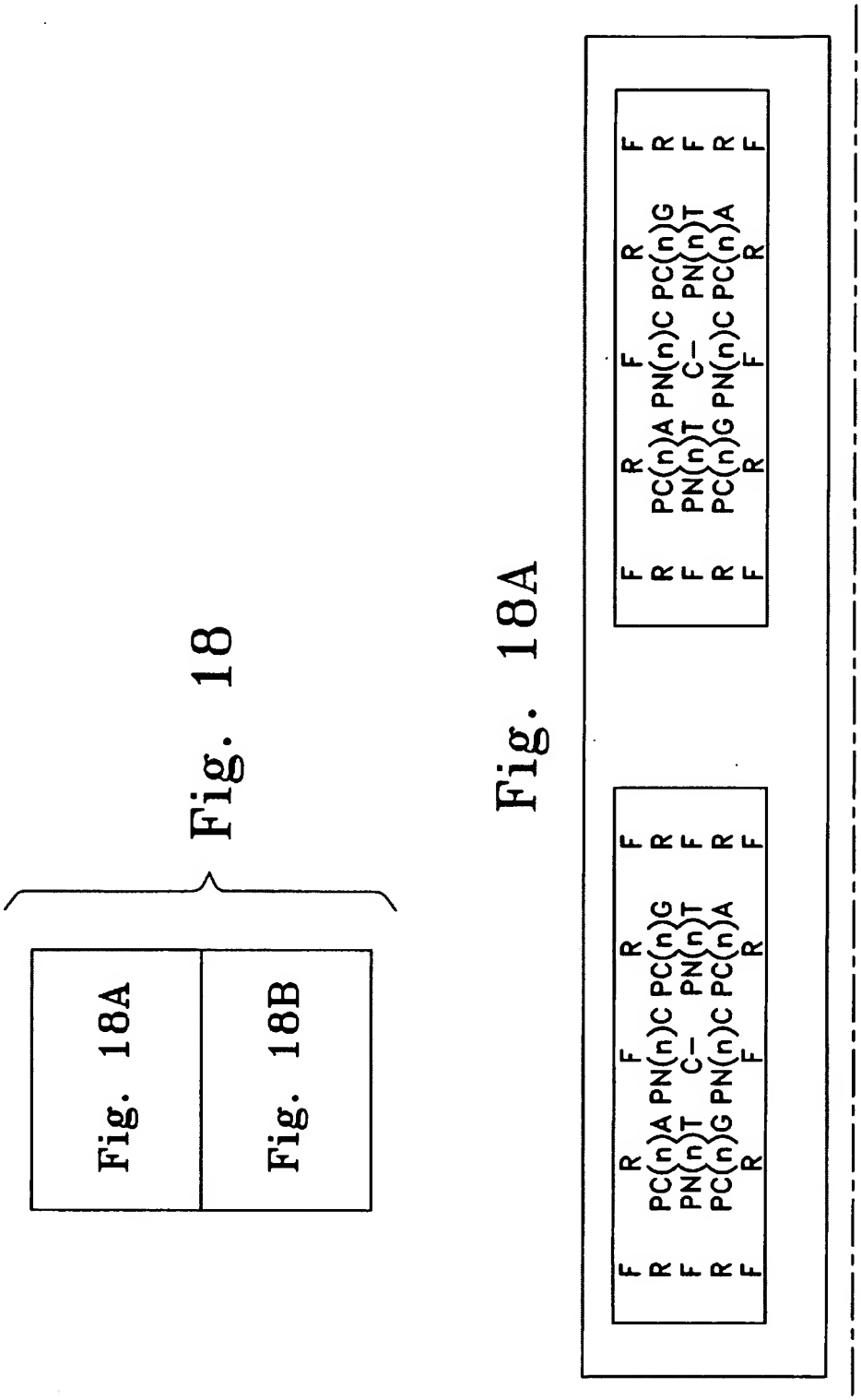


Fig. 18A

Fig. 18

Fig. 18A

Fig. 18B

Probe	Description	ID	Name	3-5 Sequence
F	Forward primer	+	CompF	AAAATAGCCTCTAAGCCCAACTC
R	Reverse primer	+	CompR	GTCAATCAACTTCTGTACTGGGC
PC(n)A	Probe coding (n) A	A		TCACATTGTAAGAAGAAACCA
PC(n)C	Probe coding (n) C	C		TCACATTGTAAGAAGAAACCC
PC(n)G	Probe coding (n) G	G		TCACATTGTAAGAAGAAACCG
PC(n)T	Probe coding (n) T	T		TCACATTGTAAGAAGAAACCT
PN(n)A	Probe non-coding (n) A	A		GAGACACCACCCACCCA
PN(n)C	Probe non-coding (n) C	C		GAGACACCACCCACCCC
PN(n)G	Probe non-coding (n) G	G		GAGACACCACCCACCCG
PN(n)T	Probe non-coding (n) T	T		GAGACACCACCCACCCT
PC(n-1)A	Probe coding (n-1) A	A		TGATCACATTGTAAGAAGAAACA
PC(n-1)C	Probe coding (n-1) C	C		TGATCACATTGTAAGAAGAAACC
PC(n-1)G	Probe coding (n-1) G	G		TGATCACATTGTAAGAAGAAACG
PC(n-1)T	Probe coding (n-1) T	T		TGATCACATTGTAAGAAGAAACT
PN(n-1)A	Probe non-coding (n-1) A	A		GGAGACACCACCCACCA
PN(n-1)C	Probe non-coding (n-1) C	C		GGAGACACCACCCACCC
PN(n-1)G	Probe non-coding (n-1) G	G		GGAGACACCACCCACCG
PN(n-1)T	Probe non-coding (n-1) T	T		GGAGACACCACCCACCT
C-	negative control	-	(ACTG)	ACTGACTGACTGACTGACTG

Fig. 18B

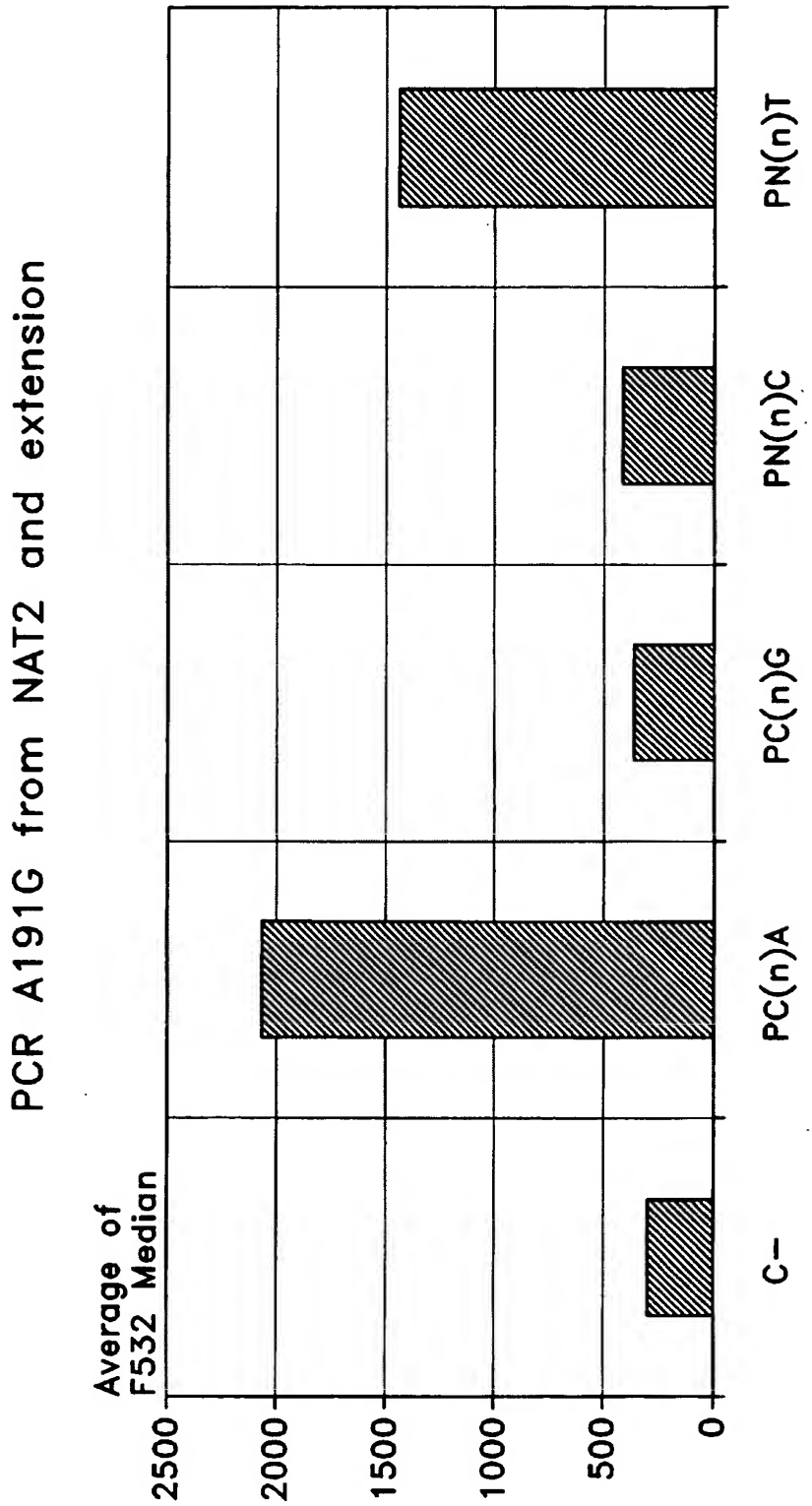

PCR A191G from NAT2 and extension
Average of F532 Median
2500
2000
1500
1000
500
0
C-
PC(n)A
PC(n)G
PN(n)C
PN(n)T

Fig. 19

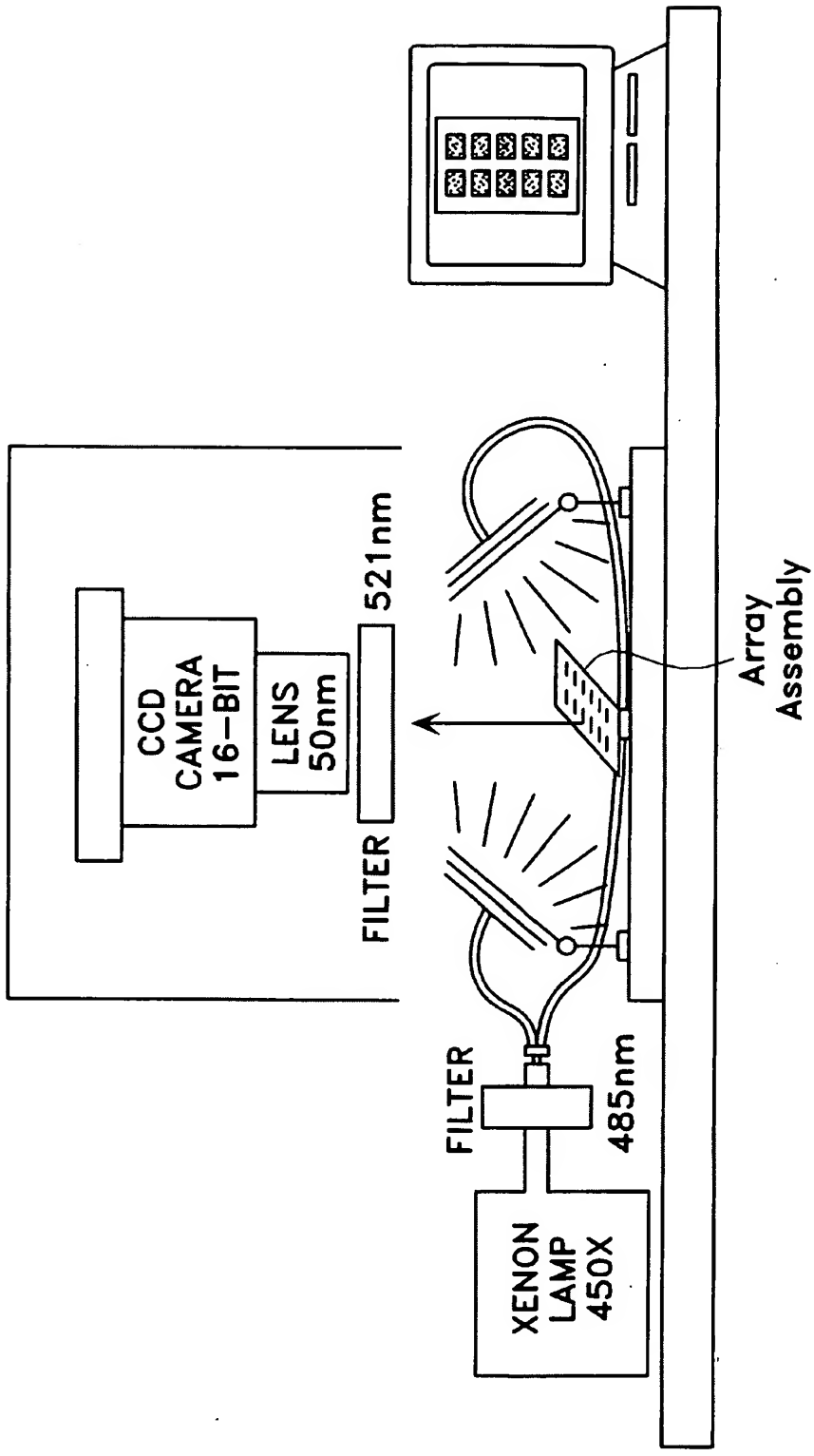

CCD CAMERA 16-BIT
LENS 50nm
521nm
FILTER
Array Assembly
FILTER
485nm
XENON LAMP 450X

Fig. 20

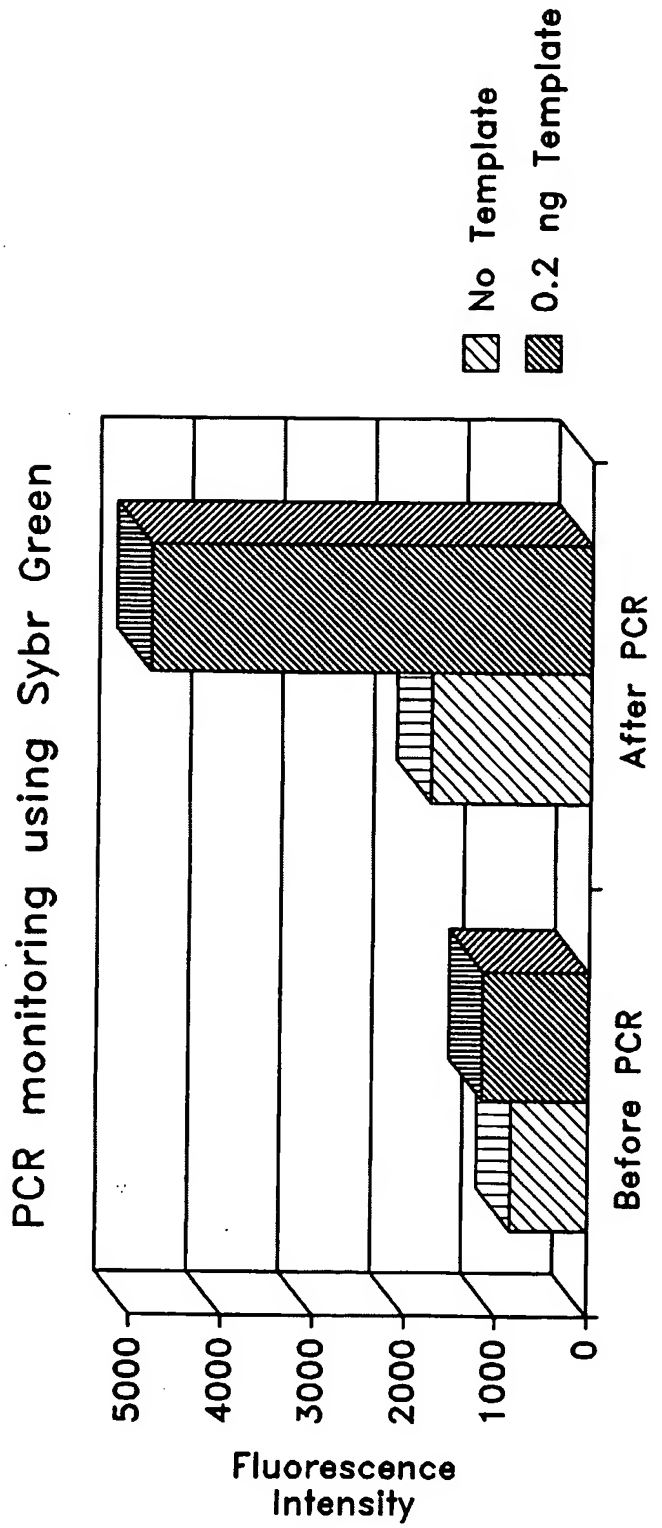

No Template
0.2 ng Template
PCR monitoring using Sybr Green
After PCR
Before PCR
5000
4000
3000
2000
1000
0
Fluorescence Intensity

Fig. 21

METHOD AND APPARATUS FOR PERFORMING LARGE NUMBERS OF REACTIONS USING ARRAY ASSEMBLY WITH RELEASABLE PRIMERS

This application claims priority to U.S. Provisional Application Ser. No. 60/158,315, filed Oct. 8, 1999 (Attorney Docket No. 05871.0010.00US00) and is a continuation U.S. Non-Provisional Application Ser. No. 09/684,736, filed Oct. 6, 2000 .

FIELD OF THE INVENTION

The present invention relates to a method and apparatus for performing a large number of reactions using array assembly. In particular, the present invention features a method and apparatus for performing a large number of chemical and biological reactions by bringing two arrays into close apposition and allowing reactants on the surfaces of two arrays to come into contact. The present invention is exemplified by performing a large number of polynucleotide amplification reactions using array assembly. In addition, the present invention features a method and apparatus for coupling the amplification of polynucleotides and the detection of sequence variations, expression levels, and functions thereof.

BACKGROUND OF THE INVENTION

Intense efforts are under way to map and sequence the human genome and the genomes of many other species. In June 2000, the Human Genome Project and Celera Genomics announced that a rough draft of the human genome had been completed. This information, however, represents only a reference sequence of the 3-billion-base human genome. The remaining task lies in the determination of sequence variations (e.g., mutations, polymorphisms, haplotypes) and sequence functions, which are important for the study, diagnosis, and treatment of human genetic diseases.

In addition to the human genome, the mouse genome is being sequenced. Genbank provides about 1.2% of the 3-billion-base mouse genome and a rough draft of the mouse genome is expected to be available by 2003 and a finished genome by 2005. The Drosophila Genome Project has also been completed recently. Thus far, genomes of more than 30 organisms have been sequenced.

Traditional nucleic acid sequencing methods include the chemical cleavage method (or the Maxam-Gilbert method) and the chain termination method (or the Sanger method) (Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989). The basic strategy for the chemical cleavage method is to specifically cleave the end-labeled DNA at only one type of nucleotide, which produces a set of labeled fragments. These labeled fragments are then separated according to their size by electrophoresis. The DNA sequence can be directly read off an autoradiogram. The chain termination method utilizes a DNA polymerase to make complementary copies of the single-stranded DNA being sequenced in the presence of a suitable primer and four deoxynucleoside triphosphates (dNTPs), of which at least one is labeled. In addition, a small amount of the 2', 3'-dideoxynucleoside triphosphate of one of the bases is added to the sequencing reaction, which generates a series of truncated chains. Each truncated chain is terminated by the dideoxy analog at positions occupied by the corresponding base, because of the absence of a 3'-OH group. Electrophoresis separates these truncated chains according to their sizes, thus indicating the positions at which dideoxy incorporation occurs and in turn the corresponding normal nucleotide. Although the efficiency of the traditional sequencing methods has been improved by automation, the use of gel electrophoresis in both methods presents a limitation on the rate of sequencing.

While the traditional chemical cleavage and chain termination sequencing methods are capable of identifying the sequence of all nucleotides in a target nucleic acid, it is quite sufficient in many cases to know the sequence identity of a single nucleotide (or a few nucleotides) at a predetermined site, i.e., the detection of known sequence variations. During the past decade, the development of array-based hybridization technology has received great attention. This high throughput method, in which hundreds to thousands of polynucleotide probes immobilized on a solid surface are hybridized to target nucleic acids to gain sequence and function information, has brought economical incentives to many applications. See, e.g., McKenzie, S., et al., *European Journal of Human Genetics* 416–429 (1998); Green et al., *Curr. Opin. in Chem. Biol.* 2:404–410 (1998), Gerhold et al., *TIBS*, 24:168–173 (1999), Young, *Cell* 102:9–15 (2000), and U.S. Pat. Nos. 5,700,637, 6,054,270, 5,837,832, 5,744,305, and 5,445,943.

DNA array-based sequencing technology generally falls into two categories. The first category is sequencing by polynucleotide hybridization. Sets of polynucleotide probes, that differ by having A, T, C, or G substituted at or near the central position, are immobilized on a solid support by in situ synthesis or by deposition of pre-synthesized polynucleotide probes. Labeled target nucleic acids containing the sequences of interest will hybridize best to perfectly matched polynucleotide probes, whereas sequence variations will alter the hybridization pattern, thereby allowing the determination of mutations and polymorphic sites (Wang, D., et al., *Science* 280:1077–1082 (1998), Lipshutz, R., et al., *Nature Genetics Supplement* 21:20–24 (1999), and Drmanac et al., *Nature Biotechnology* 16:5–58 (1998)).

Alternatively, the de novo sequencing of target nucleic acids by polynucleotide hybridization may also be accomplished. For example, an array of all possible 8-mer polynucleotide probes may be hybridized with fluorescently labeled target nucleic acids, generating large amounts of overlapping hybridization data. The reassembling of this data by computer algorithm can determine the sequence of target nucleic acids. See, e.g., Drmanac, S. et al., *Nature Biotechnology* 116:54–58 (1998), Drmanac, S. et al. *Genomics* 4:114–28 (1989), and U.S. Pat. Nos. 5,202,231, 5,492,806, 5,525,464, 5,667,972, 5,695,940, 5,972,619, 6,018,041, and 6,025,136.

The second category is sequencing by primer extension reactions (also known as minisequencing). Typically, a DNA polymerase is used specifically to extend an interrogation primer, which anneals to the nucleic acids immediately 3' of the single base substitution of interest, with a single labeled nucleoside triphosphate complementary to the single base substitution (Syvänen, *Human Mutation* 13:1–10 (1999), Syvänen et al., *Genomics* 8:684–692 (1990), Sokolov, *Nucleic Acids Res.* 18:3671 (1990), and Kuppuswami et al., *Proc. Natl. Acad. Sci. USA* 88:1143–1147 (1991)).

While methods of hybridization and primer extension-based nucleic acid sequencing have gained widespread acceptance in commercial areas, there are many limitations to the existing methods. The current methods for determining polynucleotide variations in a target nucleic acid employ discrete amplification steps and sequencing steps

(Landegren et al., *Genome Res.* 8:769–776 (1998)). Thus, additional amount of time and labor is required to separate amplification products from the amplification primers and dNTPs before the sequencing reaction. Further, it is estimated that at least about 3,000,000 single nucleotide polymorphisms (SNPs) exist in an individual's genome. As SNPs are dispersed throughout the genome, it is necessary to amplify a large number of discrete regions in the genome so that each SNP can be analyzed. Accordingly, the genetic analysis of a single individual's SNPs can require more than 3,000,000 amplification reactions be carried out and the product of each amplification reaction be analyzed. In addition, genetic analysis of a disease may require extensive genotyping of hundreds of thousands of individuals. Therefore, the number of separate amplification and sequencing reactions can be in the millions. The cost in terms of time, labor, equipment, laboratory space and reagents for carrying out discrete amplification and sequencing reactions on a large-scale is prohibitively high. Finally, the designing, optimizing and manufacturing of probe-immobilized arrays can be costly as well. For example, photolithographic synthesis of an array with N-mer polynucleotides typically requires 4×N different chrome photolithographic masks (i.e., 100 different chrome masks for a 25-mer synthesis) (Singh-Gasson et al., *Nature Biotechnol.* 17:974–978 (1999)). This leads to high cost and long synthesis time. In addition, changing probe length and base composition in photolithographic DNA synthesis means changing masks, which again leads to high redesigning cost and long turnaround time for custom arrays. There is a need in micro array field to develop rapid and inexpensive methods for large-scale sequence variation and function analysis.

The present invention features novel applications of the array technology, in which large numbers of non-unimolecular reactions are initiated and performed by array assembly. This method is capable of generating large amounts of data or products per unit time by carrying out large numbers of reactions in parallel. Furthermore, the present invention is amendable to full automation.

SUMMARY OF THE INVENTION

The present invention relates to a method and apparatus for performing a large number of reactions using array assembly. In particular, the present invention features a method and apparatus for performing a large number of chemical and biological reactions by bringing two arrays into close apposition and allowing reactants on the surfaces of two arrays to come into contact. The present invention is exemplified by performing a large number of polynucleotide amplification reactions using array assembly. In addition, the present invention features a method and apparatus for coupling the amplification of polynucleotides and the detection of sequence variations, expression levels, and functions thereof.

Any suitable solid supports (also known as arrays, chips, etc.) may be used in the present invention. These materials include glass, silicon, quartz, nylon, polystyrene, polyethylene, polypropylene, polytetrafluorethylene, metal, among others. These materials typically have a rigid or semi-rigid surface. In some embodiments, at least one surface of the material is substantially flat. Typically, at least one solid support in the assembly is derivatized to provide covalent or noncovalent attachment to chemical or biological entities. Typically, the density of derivatized sites on an array is between about 10–10,000 per cm^2, preferably below about 5,000, 1,000, 400, 200, 100, or 60 per cm^2. The area of each site may be about 1×10^{-3} to 5 mm^2, preferably less than about 2, 1, 0.5, 0.2, or 0.1 mm^2. Typically, the total number of derivatized sites on an array is between about 20–1,000,000, preferably, between about 20–500,000, 20–100,000, 20–50,000, 20–10,000, 20–5,000, 20–1,000, or 20–500. In some embodiments, an array may contain raised or depressed regions, e.g., features such as wells, raised regions, etched trenches, etc. The dimensions of these features are flexible, depending on factors, such as desirable reaction concentration, avoidance of air bubbles upon assembly, mechanical convenience and feasibility, etc. For example, the area of a well on an array is in general larger than one derivatized site on another array. The depth of a well may be less than about 1000 microns, preferably less than about 500, 200 or 100 microns. In preferred embodiments, surface tension arrays may be used. Certain reactants such as biopolymers (polynucleotides, polypeptides, etc.) are synthesized in situ to provide better yield and great flexibility.

The array assembly method provides an environment for simultaneously carrying out between about 10–500,000 reactions, preferably, above about 20, 50, 100, 200, 500, 1,000, 5,000, 10,000, 50,000, or 100,000 reactions. A myriad of chemical and biological reactions may be carried out using the instant method and apparatus. These reactions may involve cells, viruses, nucleic acids, proteins, carbohydrates, lipids, and small molecules, among others. In particular, a large number of polynucleotide amplification reactions or molecular binding reactions may be performed using array assembly. In some embodiments of the instant invention, one or more reactants may be immobilized on an array prior to array assembly. The immobilization may be covalent or non-covalent. For example, one or more reactants may be tethered to an immobilized moiety on the array. In certain embodiments of the instant invention, one or more reactants may be immobilized on an array via a releasable site, for example by tethering to an immobilized molecule with a releasable site. The immobilized reactants may be released from an array upon reacting with cleaving reagents prior to, during or after the array assembly. The release methods may include a variety of enzymatic, or non-enzymatic means, such as chemical, thermal, or photolytic treatment. In certain embodiments, detection of sequence variations or quantitation of polynucleotides may be coupled to the amplification reactions.

In one embodiment, the instant invention provides a system for performing a plurality of reactions comprising: a first solid support having a first reactant of each reaction confined to an area on the surface of said first solid support; and a second solid support having a second reactant of each reaction confined to an area on the surface of said second solid support; wherein said first and second solid supports are assembled to allow said first reactant of each reaction in contact with said second reactant of each reaction, thus providing an environment for performing said plurality of reactions in parallel.

The instant invention also provides a system for performing a plurality of reactions comprising: a first solid support having a reactant of each reaction immobilized on said first solid support; and a second solid support providing a chemical or mechanical separation of said plurality of reactions; wherein said first and second solid supports are assembled to provide an environment for performing said plurality of reactions in parallel.

The instant invention also provides a solid support for performing a plurality of polynucleotide amplification reactions wherein a releasable primer for each amplification reaction is immobilized on an area of the surface of said solid support.

The instant invention also provides a system for amplifying a plurality of target nucleic acids, comprising:

(a) a first solid support wherein

(1) the surface of said first solid support comprises a plurality of derivatized areas;

(2) a primer for each target nucleic acid or a sequence complementary to said primer is immobilized on a derivatized area of the first solid support; and

(b) a second solid support wherein said second solid support comprises a plurality of wells and each well corresponds to a primer for each target nucleic acid.

The instant invention also provides a system for amplifying a plurality of target nucleic acids, comprising:

(a) a first solid support wherein

(1) the surface of said first solid support comprises a plurality of derivatized areas;

(2) a forward primer for each target nucleic acid or a sequence complementary to said forward primer is immobilized on a derivatized area of the first solid support

(3) a reverse primer for each region of said target nucleic acid or a sequence complementary to said reverse primer is immobilized on another derivatized area of the first solid support;

(b) a second solid support wherein said second solid support comprises a plurality of wells and each well corresponds to the forward and reverse primers for each target nucleic acid.

The instant invention also provides a system for amplifying a plurality of target nucleic acids and detecting amplified products, comprising:

(a) a first solid support wherein

(1) the surface of said first solid support comprises a plurality of derivatized areas;

(2) a primer for each target nucleic acid or a sequence complementary to said primer is immobilized on a derivatized area of the first solid support;

(4) a probe comprising a subsequence, the same as or complementary to a subsequence of each target nucleic acid, is immobilized on another derivatized area of the first solid support;

(b) a second solid support wherein the surface of said second solid support comprises a plurality of reaction wells and each well corresponds the primer and the probe for each target nucleic acid.

In another embodiment, the instant invention provides a method for performing a plurality of reactions, comprising the steps of:

(a) obtaining a first solid support wherein a first reactant of each reaction is confined to an area on the surface of said first solid support;

(b) obtaining a second solid support wherein a second reactant of each reaction is confined to an area on the surface of said second solid support; and

(c) assembling said first and second solid support, wherein said first reactant on said first solid support is in contact with said second reactant on said second solid support, providing an environment for performing said plurality of reactions in parallel.

The instant invention also provides a method for performing a plurality of polynucleotide amplification reactions, comprising the steps of:

(a) obtaining a first solid support wherein a plurality of immobilized moieties are confined to a plurality of areas on the surface of said first solid support and each said immobilized moiety contains a releasable site and a primer;

(b) obtaining a second solid support wherein the surface of said second solid support contains a plurality of areas and reactants of said polynucleotide amplification reactions are confined on each area of said second solid support;

(c) assembling said first and second solid supports, wherein said reactants of said polynucleotide amplification reactions on said second solid support are in contact with said immobilized moieties on said first solid support; and

(d) releasing said primers.

The instant invention also provides a method for performing a plurality of polynucleotide amplification reactions and capturing amplification products, comprising the steps of:

(a) obtaining a first solid support wherein a plurality of immobilized moieties are confined to a plurality of areas on the surface of said first solid support and each said immobilized moiety contains a releasable site and a primer;

(b) obtaining a second solid support wherein the surface of said second solid support contains a plurality of areas and reactants of said polynucleotide amplification reactions are confined on each area of said second solid support;

(c) assembling said first and second solid supports, wherein said reactants of said polynucleotide aamplification reactions on said second solid support are in contact with said immobilized moieties on said first solid support;

(d) releasing said primers;

(e) generating amplification products of said polynucleotide amplification reactions; and

(f) capturing said amplified products by a plurality of immobilized polynucleotide probes on either said first or second solid support through hybridization.

The instant invention also provides a method for detecting a plurality of polynucleotide sequence variations, comprising the steps of:

(a) obtaining a first solid support wherein a plurality of immobilized moieties are confined to a plurality of areas on the surface of said first solid support and each said immobilized moiety contains a releasable site and a primer;

(b) obtaining a second solid support wherein the surface of said second solid support contains a plurality of areas and reactants of said polynucleotide amplification reactions are confined on each area of said second solid support;

(c) assembling said first and second solid support, wherein said reactants of said polynucleotide amplification reactions on said second solid support are in contact with said immobilized moieties on said first solid support;

(d) releasing said primers;

(e) generating amplification products of said polynucleotide amplification reactions;

(f) capturing said amplified products by a plurality of immobilized polynucleotide probes on either said first or second array through hybridization; and

(g) detecting polynucleotide sequence variations by hybridization complexes in step (f).

The instant invention also provides a method for quantitating polynucleotides in a target nucleic acid, comprising the steps of:

(a) obtaining a first solid support wherein a plurality of immobilized moieties are confined to a plurality of finite areas on the surface of said first solid support and each said immobilized moiety contains a releasable site and a primer;

(b) obtaining a second solid support wherein the surface of said second solid support contains a plurality of areas and reactants of said polynucleotide amplification reactions are confined on each area of said second solid support;

(c) assembling said first and second solid support, wherein said reactants of said polynucleotide amplification reactions on said second solid support are in contact with said immobilized moieties on said first solid support;

(d) releasing said primers;

(e) generating amplification products of said polynucleotide amplification reactions; and

(f) quantitating amplified products.

The instant invention also provides a method for detecting polynucleotide sequence variations in a target nucleic acid, comprising the steps of:

(a) obtaining a first solid support wherein a plurality of immobilized moieties are confined to a plurality of areas on the surface of said first solid support and each said immobilized moiety contains a releasable site and a primer;

(b) obtaining a second solid support wherein the surface of said second solid support contains a plurality of areas and reactants of said polynucleotide amplification reactions are confined on each areas of said second solid support;

(c) assembling said first and second solid supports, wherein said reactants of said polynucleotide amplification reactions on said second solid support are in contact with said immobilized moieties on said first solid support;

(d) releasing said primers;

(e) generating amplification products of said polynucleotide amplification reactions; and

(f) capturing said amplified products by a plurality of immobilized polynucleotide probes on either first or second array through hybridization; and

(g) detecting polynucleotide sequence variations by a polynucleotide modifying enzyme.

The instant invention also provides a method for amplifying a target nucleic acid, capturing the amplified product, and detecting a polynucleotide sequence variation in the amplified product, comprising the steps of:

(a) obtaining a first solid support wherein:

(1) the surface of said first solid support comprises a first, second, third, and fourth areas;

(2) a first chemical moiety, comprising a releasable forward primer for said target nucleic acid, is immobilized on said first area;

(3) a second chemical moiety, comprising a releasable reverse primer for said target nucleic acid, is immobilized on said second area;

(4) a first polynucleotide probe, comprising a subsequence complementary to one variant of said polynucleotide variation, is immobilized on said third area, said subsequence containing at least one interrogation position complementary to a corresponding nucleotide in said variant; and

(5) a second polynucleotide probe is immobilized to said fourth area, said second polynucleotide probe differing from said first polynucleotide probe by at least one nucleotide;

(b) obtaining a second solid support wherein the surface of said solid support comprises a reaction well and a mixture of reactants comprising a DNA polymerase, said target nucleic acid, and deoxynucleotides are placed within said reaction well;

(c) assembling said first and second solid support, wherein said mixture of reactants are in contact with said first, second, third, and fourth areas on said first solid support;

(d) releasing said releasable forward and reverse primers;

(e) generating the amplified product for said target nucleic acid;

(f) capturing the amplified product by said first or second polynucleotide probe through hybridization;

(g) disassembling said first and second solid supports;

(h) washing said first solid support;

(i) comparing the relative binding of two probes on said first solid support; and

(j) identifying said polynucleotide variation in the amplified product.

The instant invention also provides a method for amplifying a target nucleic acid, capturing the amplified product, and detecting a polynucleotide sequence variation in the amplified product, comprising the steps of:

(a) obtaining a first solid support wherein:

(1) the surface of said first array comprises a first, second, third, and fourth areas;

(2) a first chemical moiety, comprising a releasable forward primer specific for said region of said target nucleic acid, is immobilized on said first area;

(3) a second chemical moiety, comprising a releasable reverse primer specific for said region of said target nucleic acid, is immobilized on said second area;

(4) a first polynucleotide probe, comprising a subsequence complementary to one variant of said polynucleotide variation, is immobilized on said third area, said subsequence containing at least one interrogation position complementary to a corresponding nucleotide in said variant; and

(5) a second polynucleotide probe is immobilized to said fourth area, said second probe differing from said first probe by at least one nucleotide;

(b) obtaining a second solid support wherein the surface of said solid support comprises a reaction well and a mixture of reactants comprising a DNA polymerase, said target nucleic acid, and deoxynucleotides are placed within said reaction well;

(c) assembling said first and second solid support, wherein said mixture of reactants are in contact with said first, second, third, and fourth areas on said first solid support;

(d) releasing said releasable forward and reverse primers;

(e) generating the amplified product for said target nucleic acid;

(f) capturing the amplified product by said first or second polynucleotide probes through hybridization;

(g) extending said one or more hybridization complexes in step (f);

(h) disassembling said first and second solid support;

(i) washing said first solid support; and

(j) identifying said polynucleotide variation using said one or more extended products in step (g).

BRIEF DESCRIPTION OF THE FIGURES

FIG. 1 illustrates the merging of large numbers of microdroplets by bringing two arrays into close apposition.

FIG. 2 illustrates array assembly using a first array and a second array (microfabricated with wells) wherein each well of the second array faces a unit cell comprised of two or more reactant-containing areas of the first array.

FIG. 3 illustrates two arrays each containing reactants of polynucleotide amplification reactions. Large numbers of polynucleotide amplification reactions may be initiated by bringing two arrays into close apposition and merging all reactants on two arrays. In this illustration, a cleavable primer pair with different sequences (e.g., a forward primer and a reverse primer) are immobilized on two areas of the first array. Upon array assembly, the forward and reverse amplification primers are released to merge with the remaining amplification reactants in a reaction well on the second array. Each reaction well represents a different amplification using either different primer pairs or different target nucleic acid sequence.

FIGS. 4A–4C illustrate a method of amplifying a target nucleic acid, capturing the amplified products and sequencing by hybridization with captured amplified products.

FIGS. 5A–5D illustrate a method of amplifying a target nucleic acid, capturing the amplified products, and sequencing by primer extension reactions using captured amplified products.

FIGS. 6A–6D illustrates a method of amplifying a target nucleic acid, capturing the amplified products, and sequencing by primer extension, in the absence of chain terminating nucleotides, using captured amplified products.

FIG. 7 illustrates a chemical synthesis of photocleavable DMT-protected o-nitrobenzyl amidite.

FIG. 8 illustrates capillary electrophoresis analysis of polynucleotides synthesized off of a photocleavable linker.

FIG. 9 illustrates capillary electrophoresis analysis of polynucleotides synthesized off of a photocleavable linker co-migrated with polynucleotides of the same sequence from a commercial source.

FIG. 10 illustrates photolysis of in situ synthesized photocleavable polynucleotides.

FIG. 11 illustrates photolytic release of polynucleotides from the surface and subsequent hybridization of the released polynucleotides.

FIG. 12 illustrates several designs of an array micrfabricated with wells.

FIG. 13 illustrates a layout of two unit cells on a derivatized array used for detection by hybridization.

FIG. 14 illustrates a result of primer release from a hybridization complex.

FIG. 15 illustrates an example of array assembly.

FIG. 16 illustrates the result of coupling polynucleotide amplification reactions with detections by hybridization using array assembly.

FIG. 17 illustrates a comparison between primer releases from hybridization complexes vs. primers in solution.

FIG. 18 illustrate a layout of two unit cells on a derivatized array for detection by a DNA modifying enzyme.

FIG. 19 illustrates the result of coupling polynucleotide amplification reactions with detections by allele specific extension using array assembly.

FIG. 20 illustrate end-point quantitation of PCR product using array assembly.

FIG. 21 illustrates polynucleotide quantitation using a fluorescent label.

DETAILED DESCRIPTION OF THE INVENTION

Before the invention is described in detail, it is to be understood that this invention is not limited to the particular component parts or process steps of the method and apparatus described, as such parts and steps may vary. It is also to be understood that the terminology used herein is for purposes of describing particular embodiments only, and is not intended to be limiting. As used in the specification and the appended claims, the singular forms "a", "an", and "the" include plural references unless the context clearly indicates otherwise.

The present invention relates to a method and apparatus for performing a large number of reactions using array assembly. In particular, the present invention features a method and apparatus for performing a large number of chemical and biological reactions by bringing two arrays into close apposition and allowing reactants on the surfaces of two arrays to come into contact. In addition, selected reactants may be first immobilized on an array and subsequently released before, during or after the assembly with another array.

The present invention is exemplified by performing a large number of polynucleotide amplification reactions using array assembly. In particular, the present invention features a method and apparatus for coupling the amplification of polynucleotides and the detection of sequence variations, expression levels, and functions thereof.

I. Performing Large Numbers of Non-unimolecular Reactions by Array Assembly

One of skill in the art will appreciate that reactions compatible with the present method and apparatus are very broad, encompassing all non-unimolecular reactions, i.e., reactions involving two or more reactants. A myriad of chemical and biological reactions may be carried out using the instant method and apparatus. These reactions may involve cells, viruses, nucleic acids, proteins, carbohydrates, lipids, and small molecules, among others. All reactants in a non-unimolecular reaction may be separated into two groups. Each group of reactants comprises one or more reactants, but not all reactants. Therefore, each group alone is not capable of initiating the reaction. Only when one group of reactants is combined with the second group of reactants can the reaction be initiated. One group of reactants may be confined to finite areas on an array, while the other group of reactants may also confined to finite areas on another array. The combination of two groups of reactants is achieved by bringing two arrays into close apposition and allowing the merge of reactants on two arrays (FIG. 1).

While it is convenient that reactants of a reaction is divided into two groups and the reaction is triggered by assembling two reactant-containing arrays, many variations of this method are entirely within the contemplation of the instant invention. For example, reactants may be divided into more than two groups. In a reaction that requires N reactants, up to N groups of reactants may be obtained with each group containing one or more reactants, but not all N reactants. The N groups of reactants may be confined on M arrays ($2 \leqq M \leqq N$). The reaction may be initiated by successively assembling one of the N groups of reactants on one array with another one of the N groups of reactants on another array until all N reactants are combined. The order of assembling arrays may be decided by factors, such as reactant stability, reaction yields, and convenience among

others. It should be noted that arrays containing no reactants may also be assembled with those containing reactants. For example, a washing step may be achieved by assembling with an array containing only the washing solution and no reactant.

Any suitable solid supports (also known as arrays, chips, etc.) may be used in the present invention. These materials include glass, silicon, quartz, nylon, polystyrene, polyethylene, polypropylene, polytetrafluorethylene, metal, among others. These materials typically have a rigid or semi-rigid surface. In some embodiments, at least one surface of the material is substantially flat. Typically, at least one solid support in the assembly is derivatized to provide covalent or noncovalent attachment to chemical or biological entities. Typically, the density of derivatized sites on an array is between about 10–10,000 per cm^2, preferably below about 5,000, 1,000, 400, 200, 100, or 60 per cm^2. The area of each site may be about 1×10^{-3} to 5 mm^2, preferably less than about 2, 1, 0.5, 0.2, or 0.1 mm^2. Typically, the total number of derivatized sites on an array is between about 20–1,000,000, preferably, between about 20–500,000, 20–100,000, 20–50,000, 20–10,000, 20–5,000, 20–1,000, or 20–500. In some embodiments, an array may contain raised or depressed regions, e.g., features such as wells, raised regions, etched trenches, etc. The dimensions of these features are flexible, depending on factors, such as desirable reaction concentration, avoidance of air bubbles upon assembly, mechanical convenience and feasibility, etc. For example, the area of a well on an array is in general larger than one derivatized site on another array. The depth of a well may be less than about 1000 microns, preferably less than about 500, 200 or 100 microns.

There are many methods to separate reactants of one reaction from those of another and to prevent reactants of one reaction from entering another reaction before, during, or after assembly. These methods may include mechanical methods, chemical methods, or combinations thereof. For example, large numbers of reaction wells may be microfabricated on the surface of a solid support with each well providing a reaction site (FIG. 2). In some embodiments, the raised regions of reaction wells may be hydrophobic, thus keeping aqueous solution in one reaction well from entering another reaction well upon array assembly. A liquid (such as mineral oil, Fomblin®, etc.) may also be employed to separate reactants of one reaction from those of another. In some embodiments, a liquid polymer (such as Self-Seal®, nail polish, rubber cement, etc.) may be employed as a seal between two arrays or between individual reactions to prevent excess solvent evaporation. Selected areas of a solid support surface may also be chemically or photolytically treated before, during or after assembly to provide separation of reactions. For example, selected areas may be converted from hydrophobic sites to hydrophilic sites upon chemical treatment or photolysis, thus providing separation of aqueous solution of one reaction from another via a hydrophobic matrix. In addition, reaction volumes are typically small. Microdroplets are not substantially affected by gravity and their local movements are limited. Array assembly may also be accomplished in the absence of microdroplets. For example, the reactant-containing solutions on arrays may be dried or frozen before or during the array assembly.

In one embodiment of the instant invention, polynucleotide amplification reactions may be performed by bring at least two arrays into close apposition and allowing reactants of polynucleotide amplification reactions on two arrays to come in contact. All reactants in a polynucleotide amplification reaction may be separated into at least two groups. Each group of reactants comprises one or more reactants, but not all reactants. Therefore, each group alone is not capable of initiating a polynucleotide amplification reaction. Only when one group of reactants is combined with the second group of reactants can the reaction be initiated. In preferred embodiments of the instant invention, one group of reactants may be located on one or more areas on an array, while the other group of reactants may be located on one or more areas on another array. The combination of two groups of reactants is achieved by bringing two arrays into close apposition and allowing the contact of reactants on two arrays. For example, a typical PCR amplification reaction involves several reactants, including a target nucleic acid, deoxynucleotide triphosphates, two primers, a DNA polymerase, and a buffer. One group of PCR reactants may comprise two primers. The second group of reactants may comprise the remaining reactants, including a target nucleic acid, deoxynucleotide triphosphates, a DNA polymerase, and a buffer. Neither the first nor the second group of reactants alone can start the PCR amplification reaction, because the requisite reactants are missing in each group. Only when the first group is combined with the second group, i.e. when all requisite reactants are present, can the PCR reaction be initiated and performed (FIG. 3). Of course, reactants may be grouped together in many different ways. For example, only one primer may be included in the first group and the second group of reactants may comprise the remaining reactants, including a target nucleic acid, deoxynucleotide triphosphates, the other primer, a DNA polymerase, and a buffer. It is also possible that selected reactants may be present in both groups of reactants. For instance, a buffer may be included in both the first group of reactants and the second group of reactants.

The large numbers of polynucleotide amplification reactions performed between two arrays may use similar or different target nucleic acid sequences. In some embodiments, the same target nucleic acid may be present in each reaction. But different primer sequences specific for different regions of the target sequence are immobilized on one array. Therefore, each polynucleotide amplification reaction represents the amplification of a different region of the target sequence. In other embodiments, different target nucleic acid sequence may be present in each reaction. But similar set of primers specific for the same region of the target sequence are immobilized on the first array. Each polynucleotide amplification reaction thus represents amplification of the same region of target sequences from different sources. Of course, a combination of both embodiments is applicable using the instant invention.

The array assembly method provides an environment for simultaneously carrying out between about 10–500,000 reactions, preferably, above about 20, 50, 100, 200, 500, 1,000, 5,000, 10,000, 50,000, or 100,000 reactions. The advantages of the present method and apparatus are many folds. First, multiplexing gives simultaneous results of many reactions and generates large amounts of information per unit time. Second, performing large numbers of similar reactions in parallel increases the accuracy of comparative analysis by eliminating factors such as array-to-array variations, differences in reaction conditions, among others.

II. Immobilization of Reactants on Arrays

In some embodiments of the instant invention, one or more reactants may be immobilized on an array prior to array assembly. The immobilization may be covalent or non-covalent. For example, one or more reactants may be tethered to an immobilized moiety on the array. In certain

embodiments of the instant invention, one or more reactants may be immobilized on an array via a releasable site, for example by tethering to an immobilized molecule with a releasable site. The immobilized reactants may be released from an array upon reacting with cleaving reagents prior to, during or after the array assembly. The release methods may include a variety of enzymatic, or non-enzymatic means, such as chemical, thermal, or photolytic treatment.

For example, in performing large numbers of polynucleotide amplification reactions, at least one primer may be tethered to an immobilized moiety with a releasable site on the surface of a first array. The remaining reactants of a polynucleotide amplification reaction may be confined to finite areas on the surfaces of a second array, e.g., in reaction wells designed to house reactants of a polynucleotide amplification reaction. If two arrays with matched primer-containing regions and reaction wells are brought into close apposition, reactants on two arrays may merge. The primers may be released before, during or after the merging of reactants, thereby triggering reactions by mixing the requisite primers with the remaining reactants in the reaction well. Therefore, in a single physical operation, i.e., assembling two reactant-containing arrays, hundreds of thousands of separate reactions are initiated and performed.

As an example of primer release, primers may be initially hybridized to array-immobilized polynucleotides and subsequently released by strand separation from the array-immobilized polynucleotides upon array assembly. In another example of primer release, one or more primers of polynucleotide amplification reactions may be covalently immobilized on an array via a cleavable site and released before, during, or after array assembly. For example, an array-immobilized moiety may contain a cleavable site and a primer sequence. The primer sequence may be released via selective cleavage of the cleavable sites before, during, or after assembly. Typically, the immobilized moiety may itself be a polynucleotide which contains one or more cleavable sites and one or more primer polynucleotides. A cleavable site may be introduced in an immobilized moiety during in situ synthesis. Alternatively, the immobilized moieties containing releasable sites may be prepared before they are covalently or noncovalently immobilized on the solid support.

A. Array fabrication

Any suitable solid supports (also known as arrays, chips, etc.) may be used in the present invention. These materials include glass, silicon, quartz, nylon, polystyrene, polyethylene, polypropylene, polytetrafluorethylene, metal, among others. These materials typically have a rigid or semi-rigid surface. In some embodiments, at least one surface of the material is substantially flat. In some embodiments, these materials may contain raised or depressed region, e.g., features such as wells, raised regions, etched trenches, etc. Typically, at least one solid support in the assembly is derivatized to provide covalent or noncovalent attachment to chemical or biological entities. Typically, the density of derivatized sites on an array is between about 10–10,000 per cm^2, preferably below about 5,000, 1,000, 400, 200, 100, or 60 per cm^2. The area of each site may be about 1×10^{-3} to 5 mm^2, preferably less than about 2, 1, 0.5, 0.2, or 0.1 mm^2. Typically, the total number of derivatized sites on an array is between about 20–1,000,000, preferably, between about 20–500,000, 20–100,000, 20–50,000, 20–10,000, 20–5,000, 20–1,000, or 20–500.

In preferred embodiments of the instant invention, surface tension arrays, which comprise patterned hydrophilic and hydrophobic sites, may be employed. A surface tension array may contain large numbers of hydrophilic sites against a hydrophobic matrix. A hydrophilic site typically includes free amino, hydroxyl, carboxyl, thiol, amido, halo, or sulfonate group, as well as modified forms thereof, such as activated or protected forms. A hydrophobic site is typically inert to conditions of in situ synthesis. For example, a hydrophobic site may include alkyl, fluoro group, as well as modified forms thereof, etc. In surface tension arrays, a hydrophilic site is spatially segregated from neighboring hydrophilic sites because of the hydrophobic sites between hydrophilic sites. This spatially addressable pattern enables the precise and reliable location of chemical or biological entities, such as molecules, cells, viruses, etc. The free amino, hydroxyl, carboxyl, thiol, amido, halo, or sulfonate group of the hydrophilic sites may then be covalently coupled with a linker moiety (e.g., polylysine, HEG, PEG, etc.) capable of supporting chemical and biological synthesis. The hydrophilic sites may also support non-covalent attachment to chemical or biological entities, such as molecules, cells, viruses, etc. Reagents delivered to the array are constrained by surface tension difference between hydrophilic and hydrophobic sites.

The surface tension array may also be appreciated from a thermodynamic perspective of wetting. Surface tension results from an imbalance of molecular forces in a liquid. At the surface of a liquid, the liquid molecules are attracted to each other and exert a net force pulling themselves together. High values of surface tension means that molecules tend to interact strongly. Lower values mean that molecules do not interact as strongly. Water has a very high value of surface tension because it has a high degree of hydrogen bonding. Organic molecules with polar groups such as hydroxyl, carboxyl or cyano have a slightly lower surface energy than water. Pure hydrocarbons are even lower, while fluorinated compounds are very low, because the fluorine atom does not share electrons very well so only dispersion interactions (entropy of mixing) occur.

Molecules in a liquid state experience strong intermolecular attractive forces. These cohesive forces between liquid molecules are responsible for the phenomenon known as surface tension. Molecules at the surface of a liquid droplet do not have other like molecules, and as a consequence cohere more strongly with adjacent molecules which are directly associated them.

When the attractive forces are between unlike molecules, they are described as adhesive forces. The adhesive force between a water molecule and the wall of a glass capillary (i.e., the SiOH group) is stronger than the cohesive force between two water molecules at the surface. The effect of this imbalance between adhesion and cohesion is that the meniscus will turn upward and contribute to capillary action. Conversely, for mercury, the cohesive force between two mercury atoms is stronger than the adhesive force between mercury and glass, and the meniscus turns down at the wall.

When a liquid is in contact with a solid surface, the contact angle θ may be used to quantitatively measure the extent of this interaction.

$$\gamma_{SV}-\gamma_{SL}=\gamma_{LV}\cos\theta$$

where

γ_{SV} is the surface free energy of the solid,

γ_{SL} is the interfacial free energy between the solid and the liquid,

γ_{LV} is the surface free energy of the liquid.

When a droplet of liquid is in contact with a surface which is patterned into two regions which have different surface

energies, then there is a net attraction of the liquid into the region of higher surface energy. The droplet may move, as a result of the difference in surface tension between the two regions.

In other words, polar liquids wet polar surfaces in preference to nonpolar surfaces. For a patterned array where the polar synthesis regions (hydrophilic sites) are separated by nonpolar regions (hydrophobic sites), droplets of liquid are confined to a particular synthesis site, and will not migrate to an adjacent site because of the surface tension difference imposed by the nonpolar mask.

For surface tension arrays, hydrophilic sites are derivarized sites. Typically, the density of hydrophilic sites on an array is between about 10–10,000 per cm^2, preferably below about 5,000, 1,000, 400, 200, 100, or 60 per cm^2. The area of each hydrophilic site may be about 1×10^{-3} to 5 mm^2, preferably less than about 2, 1, 0.5, 0.2, or 0.1 mm^2. Typically, the total number of hydrophilic sites on an array is between about 20–1,000,000, preferably, between about 20–500,000, 20–100,000, 20–50,000, 20–10,000, 20–5,000, 20–1,000, or 20–500.

A number of methods for fabricating surface tension arrays have been described in U.S. Pat. Nos. 5,985,551 and 5,474,796. One of such methods involves coating a solid surface with a photoresist substance and then using a generic photomask to define the array patterns by exposing them to light. The exposed surface may then be reacted with a suitable reagent to form a stable hydrophobic matrix. For example, fluoroalkylsilane or long chain alkylsilane, such as octadecylsilane, may be employed to form a hydrophobic matrix. The remaining photoresist substance may then be removed and the solid support may react with a suitable reagent, such as aminoalkyl silane or hydroxyalkyl silane, to form hydrophilic regions.

The solid support may also be first reacted with a suitable derivatizing reagent to form a hydrophobic surface. For example, the hydrophobic surface may be derivatized by vapor or liquid treatment of fluoroalkylsiloxane or alkylsilane. The hydrophobic surface may then be coated with a photoresist substance, photopatterned and developed. The exposed hydrophobic surface may be reacted with suitable derivatizing reagents to form hydrophilic sites. For example, the exposed hydrophobic surface may be removed by wet or dry etch such as oxygen plasma and then derivatized by aminoalkylsilane or hydroxylalkylsilane treatment to form hydrophilic sites. The photoresist coat may be removed to expose the underlying hydrophobic sites. The hydrophilic sites may be further functionalized, if necessary, for anchoring in situ synthesis or for depositing chemical or biological entities.

Alternatively, the solid support may be first reacted with a suitable derivatizing reagent to form a hydrophilic surface. For example, the hydrophilic surface may be derivatized by vapor or liquid treatment of aminoalkylsilane or hydroxylalkylsilane. The derivatized surface may then be coated with a photoresist substance, photopatterned, and developed. The exposed surface may be reacted with suitable derivatizing reagents to form hydrophobic sites. For example, the hydrophobic sites may be formed by fluoroalkylsiloxane or alkylsilane treatment. The photoresist coat may be removed to expose the underlying hydrophilic sites. The hydrophilic sites may be further functionalized, if necessary, for anchoring in situ synthesis or for depositing chemical or biological entities.

Variations of these procedures may also be used to fabricate a solid support surface such that solution of reactants at a derivatized site is spatially separated from solution of reactants at other derivatized sites by surface tension. Separate reactions may be carried out at each derivatized site.

Photoresist substances are readily known to those of skill in the art. For example, an optical positive photoresist substance (e.g., AZ 1350 (Novolac™ type-Hoechst Celanese™) (Novolac™ is a proprietary novolak resin, which is the reaction product of phenols with formaldehyde in an acid condensation medium)) or an E-beam positive photoresist substance (e.g., EB-9 (polymethacrylate by Hoya™)) can be used.

Suitable hydrophilic and hydrophobic derivatizing reagents are also well known in the art. Preferably, fluoroalkylsilane or alkylsilane may be employed to form a hydrophobic surface and aminoalkyl silane or hydroxyalkyl silane may be used to form hydrophilic sites. As an example, a number of siloxane derivatizing reagents are listed below:

1. Hydroxyalkyl siloxanes (Silylate surface, flnctionalize with diborane, and H_2O_2 to oxidize the alcohol)
   a. allyl trichlorochlorosilane→→3-hydroxypropyl
   b. 7-oct-1-enyl trichlorochlorosilane→→8-hydroxyoctyl
2. Diol (bis-hydroxyalkyl) siloxanes (silylate surface, and hydrolyze to diol)
   a. glycidyl trimethoxysilane→→(2,3-dihydroxypropyloxy)propyl
3. Aminoalkyl siloxanes (amines require no intermediate functionalizing step)
   a. 3-aminopropyl trimethoxysilane→3-aminopropyl
4. Dimeric secondary aminoalkyl siloxanes
   a. bis (3-trimethoxysilylpropyl) amine→bis (silyloxylpropyl)amine

Glass (polytetrasiloxane) is particularly suitable for surface tension arrays, because of the numerous techniques developed by the semiconductor industry using thick films (1–5 microns) of photoresists to generate masked patterns of exposed glass surfaces. After sufficient cleaning, such as by treatment with O_2 radical (e.g., using an O_2 plasma etch, ozone plasma treatment, etc.) followed by acid wash, the glass surface may be derivatized with a suitable reagent to form a hydrophilic surface. Suitable reagents may include aminoalkyl silane, hydroxyalkyl silane, among others. In particular, glass surface may be uniformly aminosilylated with an aminosilane, such as aminobutyldimethylmethoxysilane (DMABS). The derivatized surface may then be coated with a photoresist substance, soft-baked, photopatterned using a generic photomask to define the array patterns by exposing them to light, and developed. The underlying hydrophilic sites are thus exposed in the mask area and ready to be derivatized again to form hydrophobic sites, while the photoresist covering region protects the underlying hydrophilic sites from further derivatization. Suitable reagents, such as fluoroalkylsilane or long chain alkylsilane, may be employed to form hydrophobic sites. For example, the exposed hydrophilic sites may be burned out with an O_2 plasma etch. The exposed regions may then be fluorosilylated. Following the hydrophobic derivatization, the remaining photoresist can be removed, for example by dissolution in warm organic solvents such as methyl isobutyl ketone or N-methyl pyrrolidone (NMP), to expose the hydrophilic sites of the glass surface. For example, the remaining photoresist may be dissolved off with sonication in acetone and then washed off in hot NMP.

A number of organic polymers also have desirable characteristics for surface tension arrays. For example, Teflon (polytetrafluoroethylene) may be used. Patterned derivatization of this type of material may be accomplished by

reactive ion or plasma etching through a physical mask or using an electron beam, followed by reduction to surface hydroxymethyl groups. Polypropylene/polyethylene may be surface derivatized by gamma irradiation or chromic acid oxidation, and converted to hydroxy or aminomethylated surfaces. Highly crosslinked polystryene-divinylbenzene (ca. 50%) is non-swellable, and may be readily surface derivatized by chloromethlylation and subsequently converted to other functional groups. Nylon provides an initial surface of hexylamino groups, which are directly active. The hydrophobic patterning of these surfaces may be effected using the same type of solution based thin film masking techniques and gas phase derivatization as glass, or by direct photochemical patterning using o-nitrobenzylcarbonyl blocking groups. Perfluoroalkyl carboxylic and sulfonic acid derivatives are now used to provide the hydrophobic mask of the underlying surface. Subsequent to the patterning of these surfaces, suitable linker moieties may be coupled to the reactive group such as the hydroxy or amino group.

In addition to the use of photoresist in generating patterned hydrophilic and hydrophobic sites, surface tension arrays may be fabricated without the use of photoresist. For example, a solid support may be first reacted with a reagent to form hydrophilic sites. The hydrophilic sites may then be reacted in selected areas. The remaining hydrophilic sites may then be reacted with a reagent to form hydrophobic sites. The protected hydrophilic sites may then be deprotected to anchor in situ synthesis or to deposit chemical or biological entities. For example, a glass surface may be first reacted with a reagent to generate free hydroxyl or amino sites. These hydrophilic sites may be reacted with a protected nucleoside coupling reagent or a linker to protect selected hydroxyl or amino sites. A protected nucleotide coupling reagent includes, for example, a DMT-protected nucleoside phosphoramidite, DMT-protected H-phosphonate, etc. A linker may be of six or more atoms in length. The unprotected hydroxyl or amino sites may then be reacted with a reagent, for example, perfluoroalkanoyl halide, to form hydrophobic sites inert to in situ polynucleotide synthesis. The protected hydrophilic sites may be deprotected to anchor in situ polynucleotide synthesis. Variations of these procedures may also be used to fabricate a solid support surface such that solution of chemical or biological entities at a derivatized site is spatially separated from solutions of chemical or biological entities at other derivatized sites.

In addition to surface tension array, many other examples of arrays and fabrication methods are well known to those skilled in the art (e.g., Green et al., *Curr.Opin. in Chem. Biol* 2:404–410 (1998), Gerhold et al., *TIBS,* 24:168–173 (1999), U.S. Pat. Nos. 6,090,995, 6,030782, 5,700,637, 6,054,270, 5,919,626, 5,858,653, 5,837,832, 5,744,305, 5,445,934, WO099/8708 Singh-Gasson et al., *Nature Biotechnology* 17:974–978 (1999), all incorporated herein by reference). They include, for example, fiber optic arrays, microelectrode arrays, digital micromirror arrays, etc.

B. Covalent or noncovalent immobilization of reactants on arrays

One of skill in the art will appreciate that there are many ways of immobilizing chemical or biological moieties directly on an array (covalently or noncovalently), anchoring them to a linker moiety, or tethering them to an immobilized moiety. These methods are well taught in the field of solid phase synthesis and micro-arrays (Protocols for oligonucleotides and analogs; synthesis sand properties, *Methods Mol. Biol.* Vol. 20 (1993), Beier et al., *Nucleic Acids Res.* 27:1970–1–977 (1999), Joos et al., *Anal. Chem.* 247:96–101 (1997), Guschin et al., *Anal. Biochem.* 250:203–211 (1997), U.S. Pat. Nos. 5,700,642 and 5,830,655, Czarnik et al., *Accounts Chem. Rev.* 29:112–170 (1996), *Combinatorial Chemistry and Molecular Diversity in Drug Discovery*, Ed. Kerwin J. F. and Gordon, E. M., John Wiley & Son, New York (1997); Kahn et al., *Modern Methods in Carbohydrate Synthesis*, Harwood Academic, Amsterdam (1996), Green et al., *Curr.Opin. in Chem. Biol.* 2:404–410 (1998), Gerhold et al., *TIBS,* 24:168–173 (1999), U.S. Pat. Nos. 6,0901995, 6,083,763, 6,030,782, 5,700,637, 6,054,270, 5,919,626, 5,858,653, 5,837,832, 5,744,305, 5,445,934, WO99/58708, DeRisi, J., et al., *Science* 278:680–686 (1997), Lockhart et al., *Nature* 405:827–836 (2000), Roberts et al., *Science* 287:873–880 (2000), Hughes et al., *Nature Genetics* 25:333–337 (2000), Hughes et al., *Cell* 102:109–126 (2000), Duggan, et al., *Nature Genetics Supplement* 21:10–14 (1999), and Singh-Gasson et al., *Nature Biotechnology* 17:974–978 (1999), and all incorporated herein by reference). Exemplary chemical moieties for immobilization attachment to solid support include carbamate, ester, amide, thiolester, (N)-functionalized thiourea, functionalized maleimide, amino, disulfide, amide, hydrazone, streptavidin or avidin/biotin, and gold-sulfide, among others. The immobilization methods generally fall into one of the two categories: spotting of presynthesized reactants and in situ synthesis of reactants.

In the first category, preprepared reactants are deposited onto known finite areas on an array. For example, traditional solid phase polynucleotide synthesis on controlled-pore glass (CPG) may also be employed and then simply printing presynthesized polynucleotides onto the array using direct touch or fine micropipetting. Polynucleotides may be synthesized on an automated DNA synthesizer, for example, on an Applied Biosystems synthesizer using 5-dimethoxytritylnucleoside β-cyanoethyl phosphoramidites. Synthesis of relatively long polynucleotide sequences may be achieved by PCR-based and/or enzymatic methods for economical advantages. Polynucleotides may be purified by gel electrophoresis, HPLC, or other suitable methods known in the art before spotted or deposited on the solid support. Typical non-covalent linkages may include electrostatic interactions, ligand-protein interactions (e.g., biotin/streptavidin or avidin interaction), and base-specific hydrogen bonding (e.g., complementary base pairs), among others. Solid supports may be overlaid with a positively charges coating, such as amino silane or polylysine and presynthesized polynucleotides are then printed directly onto the solid surface.

In addition, presynthesized compounds, such as polypeptides, polynucleotides, polysaccharides, or small molecule libraries, may be obtained commercially, such as from Integrated DNA Technologies, ArQule Chembridge, and Combichem. These presynthesized compounds may be covalently or non-covalently attached to the array surface.

Methods are also available to immobilize cells or proteins on solid supports (Mrksich et al., *Ann. Rev. Biophys. Biomol. Struct.* 25:55–78 (1996), U.S. Pat. Nos. 5,989,835 and 6,103,479, WO 97/45730, WO98/38490, WO00/50872, WO00/26408, WO00/17643, WO00/17624, WO0003246). In particular, Mrkisch et al. (*Proc. Natl. Acad. Sci. USA* 93:10775–8 (1996), all incorporated herein by reference) describe a method for attaching cells on gold with self-assembled monolayers of alkanethiolates. Singhvi et al. (*Science* 264:696–698 (1994)) describe a method for placing cells on predetermined locations of an array and controlling cell shape.

Printing may be accomplished by direct surface contact between the printing reagents and a delivery mechanism.

The delivery mechanism may contain the use of tweezers, pins or capillaries, among others that serve to transfer reactants or reagents to the surface. A variation of this simple printing approach is the use of controlled electric fields to immobilize prefabricated charged reactants to microelectrodes on the array (e.g. U.S. Pat. No. 5,929,208 and WO 99/06593, both incorporated herein by reference). For example, biotinylated polynucleotide probes may be directed to individual spots by polarizing the charge at that spot and then anchored in place via a steptavidin-containing permeation layer that covers the surface (Sosnowski et al., *Proc. Natl. Acad. Sci.* 94:1119–1123 (1997) and Edman et al., *Nucleic Acid. Res.* 25:4907–4914 (1997), both incorporated herein by reference). Print may also be accomplished without the direct contact with the solid support, such as using the "drop-on-demand" method described below. Some of the advantages of spotting technologies include ease of prototyping and therefore rapid implementation, low cost and versatility.

In the second category, reactants are prepared by in situ synthesis on the array. In situ synthesis may be performed on derivatized sites. Derivatized sites thus become in situ synthesis sites. Typically, the density of in situ synthesis sites on an array is between about 10–10,000 per cm^2, preferably below about 5,000, 1,000, 400, 200, 100, or 60 per cm^2. The area of each in situ synthesis site may be about 1×10^{-3} to 5 mm^2, preferably less than about 2, 1, 0.5, 0.2, or 0.1 mm^2. Typically, the total number of in situ synthesis sites on an array is between about 20–1,000,000, preferably, between about 20–500,000, 20–100,000, 20–50,000, 20–10,000, 20–5,000, 20–1,000, or 20–500.

Small molecules may be prepared on the array using standard solid phase organic synthesis methods. Biopolymers such as polynucleotides, polypeptides, and polysaccharides may be synthesized from biochemical building blocks such as nucleotides, amino acids, and monosaccharides in a step-wise fashion. With each round of synthesis, these building blocks are added to growing chains until the desired sequence and length are achieved in each spot. In general, in situ biopolymer synthesis on an array may be achieved by two general approaches. First, photolithography may be used to fabricate biopolymers on the array. In the case of polynucleotides, a mercury lamp may be shone through a photolithograhic mask onto the array surface, which removes a photoactive group, resulting in a 5' hydroxy group capable of reacting with another nucleoside. The mask therefore predetermines which nucleotides are activated. Successive rounds of deprotection and chemistry result in polynucleotides with increasing length. This method is disclosed in, e.g., U.S. Pat. Nos. 5,143,854, 5,489,678, 5,412,087, 5,744,305, 5,889,165, and 5,571,639, all incorporated herein by reference.

The second approach is the "drop-on-demand" method, which uses technology analogous to that employed in ink-jet printers (U.S. Pat. Nos. 5,985,551, 5,474,796, 5,700,637, 6,054,270, 6,028,189, 5,927,547, WO 98/41531, Blanchard et al., *Biosensors and Bioelectronics* 11:687–690 (1996), Schena et al., *TIBTECH* 16:301–306 (1998), Green et al., *Curr. Opin. Chem. Biol.* 2:404–410 (1998), and Singh-Gasson, et al., *Nat. Biotech.* 17:974–978 (1999), all incorporated herein by reference). This approach typically utilizes piezoelectric or other forms of propulsion to transfer reagents from miniature nozzles to solid surfaces. For example, the printer head travels across the array, and at each spot, electric field contracts, forcing a microdroplet of reagents onto the array surface. Following washing and deprotection, the next cycle of biopolymer synthesis is carried out. In the case of polynucleotides, the step yields in piezoelectric printing method typically equal to, and even exceed, traditional CPG polynucleotide synthesis. The drop-on-demand technology allows high-density gridding of virtually any reagents of interest. It is also easier using this method to take advantage of the extensive chemistries already developed for biopolymer and small molecule synthesis, for example, flexibility in sequence designs, synthesis of polynucleotide analogs, synthesis in the 5'–3' direction, etc. Because ink jet technology does not require direct surface contact, piezoelectric delivery is amendable to very high throughput. An average step yields near or above about 98% in in situ polynucleotide synthesis may be obtained. Similar methods of reagent delivery using a tip of a spring probe are described in WO 99/05308, incorporated herein by reference.

In preferred embodiments, a piezoelectric pump may be used to add reagents to the in situ synthesis of primer-containing moieties on arrays. The design, construction, and mechanism of a piezoelctric pump are described in U.S. Pat. No. 4,747,796. The piezoelectric pump may deliver minute droplets of liquid to a surface in a very precise manner. The pump design is similar to the pumps used in ink jet printing. The picopump is capable of producing 50 micron or 65 picoliter droplets at up to 10,000 Hz and can accurately hit a 250 micron target at a distance of 2 cm.

Methods for solid phase synthesis of large combinatorial peptide libraries have been described in the literature (Merrifield, *Science* 232:342–347 (1986), Atherton et al., *Solid Phase Peptide Synthesis*, IRL press, London (1989), Albericio et al., *Methods Enzymol.* 289:313–316 (1997), and U.S. Pat. Nos. 5,614,608 and 5,679,773, all incorporated herein by reference). Typically, a growing polypeptide chain is covalently anchored to a solid support and amino acids are added to the support-bound growing chain in a stepwise fashion. In order to prevent unwanted polymerization of the monomeric amino acid under the reaction conditions, protection of the N-terminus of the amino acid and α-amino group using blocking groups, such as tert-butyloxycarbonyl (Boc), fluorenylmethyloxycarbonyl (Fmoc) and the like, is necessary. After the monomer is coupled to the end of the polypeptide, the N-terminal protecting group is removed, and another amino acid is coupled to the chain. This cycle of coupling and deprotecting is continued for each amino acid until the desired length is reached. Photoremovable protecting group may be used to allow removal of selected portion of the solid support, via patterned irradiation, during the deprotection cycle of the solid phase synthesis (Fodor, et al., *Science* 251:767–773 (1991) and U.S. Pat. Nos. 5,143, 854, 5,489,678, and 5,744,305, all incorporated herein by reference.) This selectively allows spatial control of the synthesis and the next amino acid is coupled to the irradiated areas. In addition to standard solid phase peptide synthesis, PCT publication WO 99/06834, incorporated herein by reference, describes a method for immobilizing a diverse population of antibodies to a solid support.

Libraries of carbohydrate compounds may also be prepared on a solid support (Ito et al., *Curr. Opin. Chem. Biol.* 2:701–708 (1998), incorporated herein by reference). In solid-phase polysaccharide synthesis, elongation of a carbohydrate compounds generally consists of two steps: coupling of the glycosyl acceptor with the glycosyl donor, and selective deprotection of a temporary protecting group to liberate the free hydroxyl group that will be subjected to the next coupling with a glycosyl donor. At the final stage of polysaccharide synthesis, all protecting groups are removed. The first carbohydrate residue is typically attached to the

solid support via a linker molecule and the residual hydroxyl groups are capped after each step. Stable nonclassical glycosyl donors may be utilized and they may be activated under specific conditions (Toshima et al., *Chem. Rev.* 93:1503–1531 (1993), incorporated herein by reference). Sulfoxide methods developped by Kahn et al. may be used for stereoselective glycosylation of alcohol group (Yan et al., *J. Am. Chem. Soc.* 116:6953–6954 (1994), incorporated herein by reference). Trichoroacetimidate, thioglycoside, n-pentenyl glycoside are also amendable to solid phase synthesis (Rademann et al., *J. Org. Chem.* 62:3650–3653 (1997), Heckel et al., *Synlett* 171–173 (1998), Nicolaou et al., *J. Am. Chem. Soc.* 119:449–450 (1998), Rodebaugh et al., *J. Org. Chem.* 62:5660–5661 (1997), Danishefsky et al., *Science* 260:1307–1309 (1993); and Zheng et al., *J. Org. Chem.* 63:1126–1130 (1998), all incorporated herein by reference). In addition, PCT publication 98/22487, incorporated herein by reference, describes methods for synthesizing very large collections of diverse thiosaccharide derivatives attached to a solid support. U.S. Pat. No. 5,846,943 and PCT publication WO 98/21221, incorporated herein by reference, describe novel solid support matrices having toxin-binding polysaccharide covalently attached to a solid support through a linker arm. In addition to chemical synthesis, enzymatic synthesis of polysaccharides, such as glycosyltransferase-catalyzed glycosylation, has also been proved feasible (Shuster et al., *J. Am. Chem. Soc.* 116:1135–1136 (1994), Yamada et al., *Tetrahedron Lett.* 36:9493–9496 (1995), and Blixt et al., *J. Org. Chem.* 63:2705–2710 (1998), all incorporated herein by reference).

Chemical synthesis of glycopeptide may also be carried out on a solid support (Meldal et al. *Curr. Opin. Chem. Biol.* 1:552–563 (1997) and Kihlberg et al., *Methods Enzymol.* 289:221–245 (1997), both incorporated herein by reference). Frequently, glycosylated amino acids are used as building blocks (Gururaja et al., *Lett Pept. Sci.* 3:79–88 (1996); Mcdevitt et al. *J. Am. Chem. Soc.* 118:3818–3828 (1996), and Paulsen et al., *J. Chem. Perkin Trans* 1:281–293 (1997), all incorporated herein by reference).

The synthesis of small organic compounds on a solid support is also well known in the art of solid phase organic synthesis (Gordon et al., *J. Med. Chem.* 37:1385–1401 (1994); Lowe, *Acc. Chem. Res.* 24:309–317 (1995), Fruchtel, *Angew. Chem. Int. Ed. Engl.* 35:17–42 (1996), Hermkens et al., *Tetrahedron* 52:4527–4554 (1996), Thompson et al., *Chem. Rev.* 96:555–600 (1996), and Andres, et al., *Curr. Opin. Chem. Biol.* 2:353–362 (1998), all incorporated herein by reference). In particular, PCT publication WO99/09073, incorporated by reference, describes methods of carrying out organic chemistry on solid supports comprising derivatized functionalities and methods for synthesizing compounds comprising amine group or N-containing heterocycles using functionalized solid support. U.S. Pat. No. 5,545,568, incorporated herein by reference, describes a general methodology for synthesizing combinatorial libraries of various nonpolymeric compounds on solid supports, such as benzodiazepine, prostaglandins, β-turn mimetics and glycerol-derived drugs. PCT publication WO 97/35198, incorporated herein by reference, describes methods for synthesizing spatially-dispersed and positionally-encoded combinatorial chemistry libraries of oligomers. The position of each solid support in each array determines the exact identity of the oligomers. This method is very useful for the synthesis of a peptide library and a non-peptide, low molecular weight organic compound libraries. PCT publication WO 98/46247, incorporated herein by reference, describes a method for immobilizing immunosupressive agent, such as cyclosporin analogs on a solid support. PCT publication WO 99/21957, incorporated herein by reference, discloses methods for generating libraries of organometallic catalysts on solid support.

C. Immobilization of reactants with releasable sites

In some embodiments of the instant invention, reactants may be first immobilized on an array and subsequently released prior to, during, or after array assembly. For example, reactants may be delivered via selective cleavage of a cleavable site on an array-immobilized moiety. A releasable site may be introduced in immobilized moieties using standard nucleic acid, peptide, carbohydrate, lipid, or organic chemistry known to one of skill in the art. The release methods may include a variety of enzymatic, or non-enzymatic means, such as chemical, thermal, or photolytic treatment.

In the case of polynucleotides, the synthesis of many modified polynucleotides containing cleavable sites is well known in the art of polynucleotide synthesis (Verma et al., *Annu. Rev. Biochem.* 67:99–134 (1998), Venkatesan et al. *J. of Org. Chem.*, 61:525–529 (1996), Kahl et al., *J. of Org. Chem.*, 64:507–510 (1999), and Kahl et al., *J. of Org. Chem.* 63:4870–4871 (1998), and U.S. Pat. Nos. 5,739,386, 5,700, 642 and 5,830,655, all incorporated herein by reference). The cleavage methods may include a variety of enzymatic, or non-enzymatic means, such as chemical, thermal, photolytic cleavage, or a combination thereof. The cleavable site may be cleaved prior to, during, or after assembling two arrays. The cleavable site may be located along the polynucleotide backbone, for example, a modified 3'–5' internucleotide linkage in place of one of the phosphodiester groups, such as ribose, dialkoxysilane, phosphorothioate, and phosphoramidate internucleotide linkage. The cleavable polynucleotide analogs may also include a substituent on or replacement of one of the bases or sugars, such as 7-deazaguanosine, 5-methylcytosine, inosine, uridine, and the like. Typically, primers cleaved from the immobilized moieties are capable of hydrogen bonding in a sequence-specific manner and are capable of extending polynucleotide synthesis in amplification reactions. Preferably, the primer polynucleotides cleaved from the immobilized moieties contain a free 3'-OH end. The free 3'-OH end may also be obtained by chemical or enzymatic treatment, following the release of primer polynucleotides.

In some instances, the single stranded immobilized polynucleotides may be converted to a double-stranded immobilized polynucleotides, e.g., to load releasable primers, to incorporate double-stranded enzyme recognition site. There are many ways to prepare double-stranded polynucleotide arrays. One method is simply adding polynuleotides containing complementary sequences or subsequences with respect to the array-immobilized polynucleotides. Another method of preparation is by incorporating hairpin domain in the single-stranded polynucleotides known to those skilled in the art. Another method of preparation is simply using primers, polymerase, and dNTPs to make double stranded polynucleotide array. Another method is by hybridizing the single-stranded immobilized polynucleotide with a double-stranded polynucleotide containing a complementary single-stranded end, followed by treatment with DNA ligase, which results in double-stranded polynucleotides. This method is described in DeRisi et al., *Science* 278:680–686 (1997) and Braun et al., *Nature* 391:775–778 (1998), both incorporated herein by reference. Another method of preparing double-stranded polynucleotide arrays by synthesizing a constant sequence at every position on an array and then annealing and enzymatically extending a complementary primer is

described in PCT publication WO 99/07888 and Bulyk et al. *Nature Biotechnology,* 17:573–577 (1999), both incorporated herein by reference.

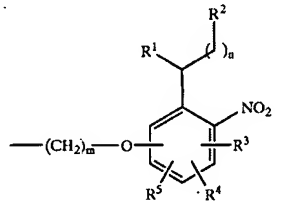
In one embodiment, cleavable sites contained within the modified polynucleotide may include chemically cleavable groups, such as dialkoxysilane, 3'-(S)-phosphorothioate, 5'-(S)-phosphorothioate, 3'-(N)-phosphoramidate, 5'-(N)-phosphoramidate, and ribose. Synthesis and cleavage conditions of chemically cleavable polynucleotides are described in U.S. Pat. Nos. 5,700,642 and 5,830,655, both incorporated herein by reference. For example, depending upon the choice of cleavable site to be introduced, either a functionalized nucleoside or a modified nucleoside dimer may be first prepared, and then selectively introduced into a growing polynucleotide fragment during the course of polynucleotide synthesis. Selective cleavage of the dialkoxysilane may be effected by treatment with fluoride ion. Phosphorothioate internucleotide linkage may be selectively cleaved under mild oxidative conditions. Selective cleavage of the phosphoramidate bond may be carried out under mild acid conditions, such as 80% acetic acid. Selective cleavage of ribose may be carried out by treatment with dilute ammonium hydroxide. Preferably, the primer polynucleotides cleaved from the immobilized moieties contain a free 3'-OH end. The free 3'-OH end may also be obtained by chemical or enzymatic treatment, following the release of primer polynucleotides.

In another embodiment, cleavable sites contained within the modified polynucleotide may include nucleotides cleavable by an enzyme such as nucleases, glycosylases, among others. A wide range of polynucleotide bases may be removed by DNA glycosylases, which cleaves the N-glycosylic bond between the base and deoxyribose, thus leaving an abasic site (for a recent review, see Krokan et. al., *Biochem. J.* 325:1–16 (1997), incorporated herein by reference). The abasic site in a polynucleotide may then be cleaved by Endonuclease IV, leaving a free 3'-OH end. Suitable DNA glycosylases may include uracil-DNA glycosylases, G/T(U) mismatch DNA glycosylases, alkylbase-DNA glycosylases, 5-methylcytosine DNA glycosylases, adenine-specific mismatch-DNA glycosylases, oxidized pyrimidine-specific DNA glycosylases, oxidized purine-specific DNA glycosylases, EndoVIII, EndoIX, hydroxymethyl DNA glycosylases, formyluracil-DNA glycosylases, pyrimidine-dimer DNA glycosylases, among others. Cleavable base analogs that are readily available synthetically are preferred and modified bases that do not preventing base pairing are also preferred. In preferred embodiments, a uracil may be synthetically incorporated in a polynucletide to replace a thymine, where the uracil is the cleavage site and site-specifically removed by treatment with uracil DNA glycosylase. The uracil DNA glycosylases may be from viral or plant sources. The abasic site on the polynucleotide strand may then be cleaved by *E. coli* Endonuclease IV.

In another embodiment, the cleavable site is a restriction endonuclease cleavable site, such as class IIs restriction enzymes. For example, BpmI, BsgI, BseRI, BsmFI, and FokI recognition sequence may be incorporated in the immobilized polynucleotides and subsequently cleaved to release primer polynucleotides. In another embodiment, the cleavable site may be a nucleotide or series of nucleotides capable of blocking or terminating 5' to 3' enzyme-promoted digestion by an enzyme having 5' to 3' exonuclease activity, such as T7 Gene 6 Exonuclease, Exo VIII, Rec J, and spleen phosphodiesterase II. Blocking nucleotides include peptide nucleic acids and nucleotides containing a phosphorothioate or borano-phosphate group (U.S. Pat. Nos. 5,700,642 and 5,830,655, both incorporated herein by reference).

In preferred embodiments, the cleavable site within an immobilized polynucleotide may include a photocleavable linker, such as ortho-nitrobenzyl class of photobleavable linkers. Synthesis and cleavage conditions of certain photolabile polynucleotides on solid support are described in Rich et al., *J.C.S. Chem. Comm.* 610–611 (1973), Venkatesan et al. *J. of Org. Chem.* 61:525–529 (1996), Kahl et al., *J. of Org. Chem.* 64:507–510 (1999), Kahl et al., *J. of Org. Chem.* 63:4870–4871 (1998), Greenberg et al., *J. of Org. Chem.* 59:746–753 (1994), McMinn et al., *Tetrahedron* 52:3827–3840 (1996), Greenberg, *Tetrahedron Lett.* 34:251–254 (1993), Yoo et al., *J. of Org. Chem.* 60:3358–3364 (1995), Greenberg, *Tetrahedron* 51:29–38 (1995), McMinn et al., *J. of Org. Chem.* 62:7074–7075 (1997), Holmes et al., *J. of Org. Chem.* 60:2318–2319 (1995), Holmes et al., *J. of Org. Chem.* 62:2370–2380 (1997), and U.S. Pat. No. 5,739,386 and 5,917,016, all incorporated by reference. Certain ortho-nitobenzyl-based linkers may also be obtained commercially. For example, hydroxymethyl, hydroxyethyl, and Fmoc-aminoethyl carboxylic acid linkers are available from Novabiochem (http://www.nova.ch). The DMT-protected o-nitrobenzyl on a solid support may also be obtained commercially (e.g., from Glen Research). For coupling photolabile linkers to hydrophilic/hydrophobic arrays, reaction of the photolabile linker with a primary amine on the array surface may be achieved using standard peptide coupling reagents.

In general, a surface immobilized photocleavable linker may have the formula of S-B-L where S is a solid support, B is a bond or a derivatizing group, and L is a photocleavable linking group having the formula:



where, R_1 is hydrogen, C_1–C_8 alkyl, aryl or arylalkyl; R_3, R_4 and R_5 are each independently hydrogen, C_1–C_8 alkyl, or C_1–C_8 alkoxy; R_2 is halogen, —SH, —SP, —OH,—NH_2, —OP or —NHP, wherein P is a suitable protecting or activating group; and n is 0 or 1 and m is an integer of from 1 to 10. This class of photocleavable linkers may be coupled to the synthesis of a variety of biopolymers such as polynucleotides and polypeptides.

In some embodiments of the invention, nonphotolabile analogue of the photolabile group, such as veratryl alcohol, may be doped in the synthesis of photocleavable immobilized polynucleotide, thus creating a heterogeneous mixture of photocleavable and nonphotocleavable immobilized polynucleotides at a finite area on the surface of an array. By varying the ratio of the amount of cleavable and noncleavable linker, the relative molar quantities of the cleaved and surface bound polynucleotides may be optimized for capturing the amplification products of amplification reactions, and for polynucleotide elongation of the complexes between the surface bound noncleavable polynucleotides and the captured amplification products.

For immobilized polypeptide in solid phase synthesis, simultaneous deprotection of the side-chain groups and the

cleavage of peptide from the solid support may be achieved by treatment with a hard acid, such as hydrogen fluoride (HF), trifluoromethanesulfonic acid (TFMSA), or trifluoromethanesulfonic acidtrimethylsilyl ester (TMSOTf) (Steward, *Methods in Enzymol.* 289:29–44 (1997), incorporated herein by referece). In addition to production of free polypeptides by acidolysis, a peptide amide may be produced by treatment with ammonia and amines. Polypeptide may also be reduced to alcohols by treating with $LiBH_4$ in THF. Sequence specific cleavage of polypeptides or modifications thereof may be carried out using suitable enzymes, such as proteases, known in the art. In addition, phenacyl based or ortho-nitrobenzyl based linking group may also be used as a photocleavable linker to couple peptides to solid support (Wang, *J. Org. Chem.* 41–3258 (1976), Rich et al., *J. Am. Chem. Soc.* 97:1575–1579 (1975), Hammer et al., *Int. J. Peptide Protein Res.* 36:31–45 (1990), U.S. Pat. Nos. 5,739,386 and 5,917,016). Photolysis offers a mild method of cleavage which complements traditional acidic or basic cleavage techniques (see, e.g., Lloyd-Williams et al., *Tetrahedron* 49:11065–11133 (1993)). Photocleavage of peptide ligands from solid supports for screening enzyme inhibitors has also been reported (Schullek et al., *Anal. Biochem.* 246:20–29 (1997)).

For immobilized polysaccharides, linker groups typically have enough stability to withstand glycosylation reaction conditions, while being able to undergo cleavage in mild conditions without affecting the polysaccharide backbone structure and protecting groups. Photolabile linkers, described in Nicolaou et al., supra and Rodebaugh et al., supra, may be used. A silyl ether may be removed by a fluoride ion (Danishefsky, et al., supra) and a thioglycoside-like linker may be cleaved under various thiophilic conditions (Yan et al., supra). In addition, base-labile linkers, such as 9-fluoroenylmethoxycarbonyl-type and succinate type may be cleavable with triethylamine and aqueous ammonia, respectively (Adinolfi et al., *Tetrahedron Lett.* 37:5007–5010 (1996) and Wang et al., *Chem. Lett.* 273–274 (1995), both incorporated herein by reference).

III. Carrying Out Large Numbers of Reactions Using Array Assembly

One of skill in the art will appreciate that all non-unimolecular reactions, which require two or more reactants are compatible with the present method and apparatus. In preferred embodiments of the instant invention, a myriad of chemical and biological reactions may be carried out using the instant method and apparatus. These reactions may involve cells, viruses, nucleic acids, proteins, carbohydrates, lipids, or small molecules. These reactions may be polynucleotide amplification reactions, or molecular binding interactions, such as, protein-DNA, nucleic acid hybridization, inhibitor-enzyme, receptor-antagonist, drug-DNA, antibody-antigen, toxin-carbohydrate, receptor-glycoprotein binding reactions, among others.

A. Polynucleotide amplification reactions

A large variety of polynucleotide amplification reactions known to those skilled in the art may be suitable for the instant invention. The most common form of polynucleotide amplification reaction, such as a PCR reaction, is typically carried out by placing a mixture of target nucleic acid sequence, deoxynucleotide triphosphates, buffer, two primers, and DNA polymerase in a thermocycler which cycles between temperatures for denaturation, annealing, and extension (*PCR Technology: Principles and Applications for DNA Amplification* (ed. H. A. Erlich, Freeman Press, NY, N.Y., 1992), *PCR Protocols: A Guide to Methods and Applications* (eds. Innis, et al., Academic Press, San Diego, Calif., 1990), Mattila et al., *Nucleic Acids Res.* 19:4967 (1991), Eckert et al., *PCR Methods and Applications* 1, 17 (1991), *PCR A Practical Approach and PCR2 A Practional Approach (eds. McPherson et al., Oxford University Press, Oxford,* 1991 and 1995), all incorporated by reference). The selection of primers defines the region to be amplified. The polymerase used to direct the nucleotide synthesis may include, for example, *E. coli* DNA polymerase I, Klenow fragment of *E. coli* DNA polymerase, polymerase muteins, heat-stable enzymes, such as Taq polymerase, Vent polymerase, and the like. For a reverse-transcriptase-PCR, enzyme such as reverse transcriptase may be used.

With the aid of arrays, large numbers of PCR reactions may be performed in parallel by confining selected PCR reactants to definite areas on two arrays, bringing two arrays into apposition, and allowing reactants on two arrays to merge and triggering PCR reactions. In preferred embodiments of the instant invention, One or more primers of the PCR reaction may also be first hybridized to immobilized moieties on defined areas of a first array surface and subsequently released by strand separation. One or more primers of the PCR reaction may be first immobilized on defined areas of an array surface via a cleavable site and subsequently released upon array assembly. The covalent array immobilization site may either be at the 5' end of the polynucleotide (or the 3' distal and 5' proximal orientation) or at the 3' end of the primer polynucleotide (or the 3' proximal and 5' distal orientation). In some instances, the array immobilization site may be within the polynucleotide (i.e. at a site other than the 5' or 3' end of the polynucleotide). The array-immobilized and primer-containing moieties typically have lengths ranging from about 5 to 100 nucleotides, preferably between about 10 to 50 nucleotides, more preferably between about 10 to 30 nucleotides.

The remaining reactants of the PCR reaction may be confined to defined areas or reaction wells on the surface of another array. The combination of two groups of reactants is achieved by bringing two arrays into close proximity and allowing the merge of reactants on two arrays. The primers of the PCR reaction may be released chemically, enzymatically, or photolytically before, during, or after assembling two arrays.

Following the primer release, the PCR reactions may be performed according to known protocols in the art. Typically, the temperature may be raised to separate the double-stranded target nucleic acid to form the single-stranded templates for amplification. The temperature may then be lowered to generate the primed templates for DNA polymerase. The temperature may again be raised to promote polynucleotide synthesis, and the cycle of strand separation, annealing of primers, and synthesis is repeated for about as many as 30–60 cycles.

One of skill in the art will appreciate that many other polynucleotide amplification reactions may also be carried out using the instant method and apparatus. For reviews, see Isaksson and Landegren, *Curr. Opin. Biotechnol.* 10:11 –15 (1999), Landegren, *Curr. Opin. Biotechnol.* 7:95–97 (1996), and Abramson et al., *Curr. Opin. Biotechnol.* 4:41–47 (1993), all incorporated herein by reference. Typically, amplification of either or both strand of the target nucleic acid comprises the use of one or more nucleic acid-modifying enzymes, such as a DNA polymerase, a ligase, an RNA polymerase, or an RNA-dependent reverse transcriptase. For example, polynucleotide amplification reactions may include self-sustained sequence replication (3SR) (Muller et al., *Histochem. Cell Biol.* 108:431–437 (1997),

incorporated herein by reference), nucleic acid sequence-based amplification (NASBA) (Malek et al., *Methods Mol. Biol.* 28:253–260 (1994), incorporated herein by reference), strand displacement activation (SDA) (Walker, *PCR Methods Appl.* 3:1–6 (1993), incorporated herein by reference), ligase chain reaction (LCR) (Wu et al., *Genomics* 4:560 (1989), Landegren et al., *Science* 241:1077 (1988), Abravaya et al., *Nucleic Acids Res.* 23:675–682 (1995), and Wiedmann et al., *PCR Methods Appl.* 3:S51–64 (1994), all incorporated herein by reference), and Qβ replicase system (Burg et al., *Anal. Biochem.* 230:263–272 (1995), incorporated herein by reference), among others. Depending on the target nucleic acid and its concentration, polynucleotide amplification reactions may either exhibit solution or solid phase reaction profile.

B. Coupling of polynucleotide amplification reactions and sequence variation detections

The present invention may be used to couple the amplification and detection procedures, thus providing an environment for simultaneously carrying out thousands of amplification reactions followed by simultaneous detection of thousands of amplified products on the array assembly. This coupling of amplification reactions and sequencing reactions simplifies the detection procedure and streamlines the methods of amplification and sequencing. Array assembly is a faster and cost-effective method for amplification and sequencing, which is amendable to high throughput applications.

In some embodiments of the instant invention, the amplification products generated in polynucleotide amplification reactions may be captured by immobilized polynucleotide interrogation probes or non-cleavable moieties on either the first or the second array. The capture may be accomplished by hybridizing amplification products with immobilized interrogation probes. For example, a fraction of immobilized polynucleotides may contain non-cleavable polynucleotides designed for capturing amplification products, while another fraction of immobilized polynucleotides confined in the same area of the array may contain cleavable polynucleotides designed for delivering amplification primers during the assembly of two arrays. It is also possible to immobilize non-cleavable moiety at a different area than the primer-containing cleavable immobilized moiety for capturing amplified products. Therefore, when amplification reactions are completed between two arrays and the arrays are pulled apart, each strand of the amplified products may be separated as each strand forms different hybridization complexes between non-cleavable polynucleotides remain immobilized on an array. It is also possible to carry out amplifications that result in an excess of one strand of the amplification products by varying primer concentrations. In this case, single stranded amplification products may also be captured by non-cleavable polynucleotides immobilized on an array.

After the amplified products of a target nucleic acid are captured, they may be directly or indirectly used in detecting polynucleotide sequence variations. Strategies for identification and detection of polynucleotide sequence variations are known in the art (Nelson, *Crit. Rev. Clin. Lab. Sci.* 35:369–414 (1998), Landegren et al., *Genome Res.* 8:769–776 (1998), Syvänen, *Human Mutation* 13:1–10 (1999), U.S. Pat. Nos. 6,001,567, 5,985,557, 5,888,819, 5,650,277, 5,710,028, 5,858,659, and 5,871,928, and PCT applications WO 99/27137, 98/54362, 98/56954, 98/38846, 99/14228, 98/30883, and 99/37812, all incorporated herein by reference).

In general, the identification of a sequence variation may be separated into two categories. In one category, the sequence variants may be distinguished by hybridization. This method employs immobilized sequence-specific polynucleotide probes. A single mismatch in a hybridization complex formed between amplification product and probes may cause a significant change in the signals detected. In a typical approach, a series of probes of known sequence and of similar length, each with one of the four different bases at a given location near the center position are affixed to the array surface.

In the second category, the sequence variants may be determined using polynucleotide modifying enzymes. The polynucleotide modifying enzymes include DNA polymerases, DNA ligase, nuclease, and restriction enzymes, among others. For example, in a minisequencing reaction, a DNA polymerase is used to extend an interrogation primer that anneals immediately adjacent to the nucleotide position of interest with a single labeled nucleoside triphosphate complementary to the nucleotide at the variant site.

1. Sequence variation detection by hybridization In one aspect, the present invention is suitable for determining precharacterized polynucleotide sequence variations. In other words, the genotyping is performed after the location and nature of polymorphic forms or mutations have already been determined. The sequences of known polymorphic forms and the wild-type/mutation sequences may be used to as reference sequences. For example, the two polymorphic forms of a biallelic single nucleotide polymorphism (SNP) may be used as two reference sequences. To analyze a deletion mutation, one can select the wild-type form and the deleted form as two reference sequences. In some instances, sequence variations of both the coding and noncoding strands of the target nucleic acid sequence may be determined. Therefore, both the coding and noncoding strands may be used as reference sequences for sequence variation determinations.

Although, in general, the reference sequences are from the same source as the target nucleic acid, in some instances, they may be from difference sources. For example, the sequences can be from a human or mouse. A substantial number of mutations and polymorphic forms have been reported in the published literature or may be accessible through publicly available web sites. See also, Gelfand et al., *Nucleic Acids Res.* 27:301–302 (1999) and Buetow et al., *Nat. Genet.* 21:323–325 (1999). The availability of reference sequence information allows an initial set of polynucleotide probes to be designed for the identification of the know sequence variations.

The determination of sequence variations using the present invention also includes de novo characterizing polynucleotide sequence variations. In other words, genotyping may be used to identify points of new variations and the nature of new variations. For example, by analyzing a group of individuals representing ethnic diversity among humans, the consensus or alternative alleles/haplotypes of the locus may be identified, and the frequencies in the population may be determined. Allelic variations and frequencies may also be determined for populations characterized by criteria such as geography, gender, among others. Such analysis may also be performed among different species in plants, animals, and other organisms.

A set of polynucleotide probes based on reference sequences for each sequence variation may be designed. The design of a probe set typically includes probes that are perfectly complementary to the reference sequences and span the location of each sequence variation. Perfect complementary means sequence-specific base pairing which

includes e.g., Waston-Crick base pairing or other forms of base pairing such as Hoogstein base pairing. Leading or trailing sequences flanking the segment of complementarity can also be present. In the simplest form, a pair of polynucleotide probes perfectly complementary to the two polymorphic forms of a biallelic SNP (two reference sequences) may be employed. Of course, additional related polynucleotide probes may be added to improve the accuracy of the detection. For example, for each perfectly complementary probe, there may be three corresponding probes, each with a different nucleotide at the variation position. More complex design of polynucleotide probes known to those skilled in the art may also be employed. For example, various tiling methods (e.g., sequence tiling, block tiling, 4x3 tiling, and opt-tiling) are described in WO 95/11995, WO 98/30883, WO 98/56954, EP 717113A2, and WO/99/39004, all incorporated herein by reference.

A diploid organism may be homozygous or heterozygous for a biallelic SNP. There are four possible SNP homozygotes (A/A, T/T, C/C, and G/G) and six possible heterozygotes (T/A, A/G, C/T, C/A, T/G, and C/G). When the polynucleotide probes are hybridized with a heterozygous sample in which both polymorphic forms are present, the patterns for the homozygous polymorphic forms are superimposed. Thus, the probes show distinct and characteristic hybridization patterns depending on which sequence variation is present and whether an individual is homozygous or heterozygous. FIG. 4 is an example of amplifying a target nucleic acid, capturing the amplified products and sequencing by hybridization with captured amplified products.

In addition to using array-immobilized polynucleoitdes for hybridization detection, many homogeneous hybridization detection methods may also be employed. For example, hybridization probe that form hairpin-loop structure in the absence of the correct target nucleic acid may be utilized (Tyagi et al., *Nature Biotechnol.* 16–49–53 (1998)). The formation of a hairpin-loop conformation brings the fluorophore and the quencher pair close together, thus extinguishing the donor fluorescence.

2. Sequence variation detection by a polynucleotide modifying enzyme

In some embodiments of the instant invention, the captured amplification products may be used indirectly in the subsequent sequencing reactions. For example, hybridization complexes between the captured amplification products and immobilized polynucleotides may be further extended using a suitable DNA modifying enzyme to determine the sequence of the amplified products.

In the primer extension method, nucleotide extension reaction catalyzed by the DNA polymerase is used to distinguish between the sequence variants. For example, a sequencing probe may be annealed to the amplified polynucleotides, e.g. immediately 3' of the sequence variation position. This primer may then be extended with one or more labeled nucleoside triphosphates that are complementary to the nucleotide to be detected using a DNA polymerase. The distinction between the sequence variants is based on the accuracy of the nucleotide incorporation reaction catalyzed by a DNA polymerase, not the differences in thermal stability between mismatched and perfectly matched hybridization complexes. This method allows discrimination between the homozygous and heterozygous genotypes. This method is robust and insensitive to small variations in the reaction conditions. The same reactions may be employed for detecting any nucleotide variable nucleotide irrespective of the nucleotide sequence flanking the variable site.

Typically, any DNA polymerase without proofreading activity may be used in this method in order to avoid 3'–5' degradation of the probe. DNA polymerase may include Klenow DNA polymerase, T7 DNA polymerase ("Sequenase"), thermostable polymerase, among others. Depending upon the polymerase, different conditions may be used, and different temperature ranges may be required be the hybridization and extension reactions. Any labeled nucleotide analogue that is incorporated sequence specifically by a DNA polymerase may serve as detectable group. In one embodiment, labeled extension terminators, such as dideoxoynucleotides (e.g., ddATP, ddCTP, ddGTP, ddTTP, and ddUTP), arabinoside triphosphates, may be used because they terminate the extension reaction. Therefore, enzymatic extension of the sequencing primer by one nucleotide depends on the correct base pairing of the added nucleotide to the nucleotide variation to be detected. In another embodiment, labeled deoxynucleotides may be used to extend the primer beyond one nucleotide. Therefore, enzymatic extension of the sequencing primer by more than one nucleotide depends on the correct base pairing of the sequencing primer to the nucleotide variation to be detected.

To obtain specific extension of the polynucleotide in the sequencing step, the excess of amplification reagents are removed before the sequencing reaction. A second separation step may also be used to separate the labeled extension products from the unincorporated labeled nucleotide before measurement. In addition, any size fractionating methods may be used to determine the sequence of primer extension products. These methods include gel electrophoresis, such as polyacrylamide or agarose gel electrophoresis, capillary electrophoresis, mass spectrometry, and HPLC.

In another example, the 5' nuclease activity of a variety of enzymes may also be employed to cleave and detect target-dependent cleavage structure (also known as the Invader® assay). Specific DNA and RNA sequences are detected by using structure-dependent enzymes to cleave a hybridization complex with overlapping polynucleotide probes. Details of detecting sequence variations using 5' nuclease assay have been described in WO 98/4287. Details of detecting sequence variations using structure-bridging polynucleotides have been described in WO 98/50403.

In addition, the sequence variants may be determined by using mass spectrometry, such as matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry, tandem mass spectrometry, etc. Certain examples of sequence variation detection using mass spectrometry have been described in U.S. Pat. Nos. 6,074,823, 6,043,031 and 5,691,141. Sequence variations may be distinguished by marker, hybridization pattern, or discrimination of DNA modifying enzyme (e.g., exonuclease) followed by mass spectrometry.

FIG. 5 is an example of amplifying a target nucleic acid, capturing the amplified products, and sequencing by primer extension reactions using captured amplified products. FIG. 6 is an example of amplifying a target nucleic acid, capturing the amplified products, and sequencing by primer extension, in the absence of chain terminating nucleotides, using captured amplified products.

B. Coupling of polynucleotide amplification reactions and quantitations of polynucleotides

The multiplexing method and apparatus disclosed in the instant invention may be used to quantitate nucleic acid molecules, for example measuring gene expression, gene copy number, viral load, etc.

One application is the monitoring of gene expression level and comparing of gene expression patterns following PCR amplification of target nucleic acid. Many gene-specific

polynucleotide probes derived from the 3' end of RNA transcripts may be spotted on a solid support. This array is then probed with fluorescently labeled cDNA representations of RNA pools from test and reference cells. The relative amount of transcript present in the pool is determined by the fluorescent signals generated and the level of gene expression is compared between the test and the reference cell. See, e.g., Lockhart et al., *Nature* 405:827–836 (2000), Roberts et al., *Science* 287:873–880 (2000), Hughes et al., *Nature Genetics* 25:333–337 (2000), Hughes et al., Cell 102:109–126 (2000), Duggan, D., et al., *Nature Genetics Supplement* 21:10–14 (1999), DeRisi, J., et al., *Science* 278:680–686 (1997), and U.S. Pat. Nos. 5,800,992, 5,871,928, and 6,040,138, all incorporated herein by reference.

In the instant invention, gene expression monitoring or profiling may be performed on a solid support with in situ synthesized polynucleotides. The simultaneous monitoring of the expression levels of a multiplicity of genes permits comparison of relative expression levels and identification of biological conditions (e.g., disease detection, drug screening) characterized by alterations of relative expression levels of various genes. The simultaneous monitoring of the expression levels also includes the determination of the presence or absence of genes.

Polynucleotide probes for expression monitoring may include probes each having a sequence that is complementary to a subsequence of one of the genes (or the mnRNA or the corresponding antisense cRNA). The gene intron/exon structure and the relatedness of each probe to other expressed sequences may also be considered. Polynucleotide probe set may additionally include mismatch controls, normalization probes, and expression level control probes, among others. In particular, expression level control probes are those hybridize specifically with constituively expressed genes in the biological sample, such as β-actin, the transferrin receptor gene, the GAPDH gene, and the like. They are designed to control for the overall health and metabolic activity of a biological sample.

In some embodiments, a multiplicity of genes may be monitored in real time by using reverse transcriptase PCR. Quantitation of transcription levels of multiple genes can be absolute or relative. Absolute quantification may be accomplished by inclusion of known concentration of one or more target nucleic acids such as control nucleic acids or known amounts of the target nucleic acids to be detected. The relative quantification may be accomplished by comparison of hybridization signals between two or more genes, or between two or more treatments to quantify the changes in hybridization intensity.

In addition of quantitating polynucleotides based on hybridization, a polynucleotide modifying enzyme may also employed to quantitatively measure polynucleotides (see, e.g., Lie et al., *Curr. Opin. in Biotech.* 9:43–48 (1998), incorporated herein by reference). These methods include for example quantitative PCR, competitive PCR, and 5' nuclease assay.

In a 5' nuclease assay (also known as the TaqMan™ technology), a polynucleotide probe is annealed to a target sequence located between the forward (5') and reverse (3') primer binding sites. The probe may be labelled with a reporter dye (e.g. FAM, 6-carboxyfluorescein) at the 5' end and a quencher dye (e.g. TAMRA, 6-carboxytetramethylrhodamine) in the middle, or at the 3' end, which compensates the emission spectra of the reporter dye as long as both dyes are attached to the probe. Modification of the probe with a 3'-blocking phosphate may prevent extension of the annealed probe during amplification. Cleavage of the probe by 5'–3' exonuclease activity of a thermostable polymerase (e.g., a Taq polymerase) during strand elongation releases the reporter from the probe and thus its proximity to the 3' quencher, resulting in an increase in reporter emission intensity. Thus, the fluorescence signals are detectable and can be quantified. The cycle at which the emission intensity of the sample rises above baseline (threshold cycle) is inversely proportional to the target sequence concentration. The higher the target concentration, the lower the number of amplification cycles required to detect the rise in reporter emission above baseline. Typically, an ABI PRISM™ Sequence Detection System (PE Applied Biosystems) may be used to detect the fluorescent signals. Two probes labelled with different reporter dyes (e.g., FAM, tetrachloro-6-carboxy-fluorescein [TET], hexachloro-6-carboxyfluorescein [JOE]) may also be used in the 5' nuclease assay. This two-color detection may improve sample quantitation accuracy, reduce assay sample numbers and increase sample throughput.

The 5' nuclease assay may be improved by measuring the released fluorescent emission continuously during the PCR amplification instead of an endpoint measurement (Real Time PCR quantification, Heid et al., *Genome Res.* 6:986–994 (1996); Gibson et al. *Genome Res.* 6:995–1001 (1996); and Livak et al. *PCR Methods and Applications* 4:357–362 (1995), all incorporated herein by reference). Since the exponential accumulation of the fluorescent signal directly reflects the exponential accumulation of the PCR amplification product, this reaction is monitored in real time. Real time RT-PCR may also be performed by adding reagents for RT-PCR and incubating reaction mixture at a temperature optimal for reverse transcription.

There are also many endpoint methods to detect and quantitate PCR products. These methods typically employ an intercalator or major or minor groove binder. For example, Higuchi et al. (*BioTechnology* 10:413–417 (1992) and 11:1026–1030 (1993)) describe a method for real-time PCR product quantitation by measuring the increase in ethidium bromide intensity during amplification with a charge-coupled device (CCD) camera. Ishiguro et al. (*Anal. Biochem.* 229:207–213 (1995)) describe the use of various intercalaters to quantitate PCR amplification products.

Quantitation of nucleic acids may also be achieved by diluting or dividing a solution containing the nucleic acid of interest. The polynucleotide amplification reactions may be performed on the diluted or divided reaction volumes until no subject nucleic acid molecules are amplifiable in the reaction volumes. The number of reaction volumes and the dilution factor can determine the number of the nucleic acid molecules in a solution.

If a solution (e.g. 100 ul) containing 1000 molecules of nucleic acid of interest is diluted to 10,000 ul and then divided into 1000 reaction volumes, each reaction volume (10 ul) contains about 0.1 ul of the original undiluted solution and theoretically 1 molecule of the nucleic acid of interest. If a polynucleotide amplification reaction is performed on each of the 1000 reaction volumes, the number of reaction volumes where amplification occurs is theoretically 1000. Of course, the number of reaction volumes where nucleic acid of interest is amplified is almost certainly less than 1000, because some reaction volumes may not get any molecule of nucleic acid of interest (i.e. some reaction volumes get two or more molecules by chance or clumping of molecules) and/or because the efficiency of amplification reaction is less than 100%. However, as the dilution factor increases, multiple seeding of nucleic acids in each reaction volume will decrease and more and more reaction volumes

will receive no nucleic acid of interest because of low concentration. When the number of reaction volume where nucleic acid of interest is amplified is substantially lower than the number of reaction volumes, e.g. less than half, the concentration of nucleic acid of interest may be accurately determined by counting the number of the reaction volume where the nucleic acid of interest is amplified and correcting it with the dilution factor. In addition, by carrying out this process with known standard amounts of nucleic acids, the efficiency of amplification reaction may be measured and subsequently used as a correction factor. For example, a solution of 100 ul contains unknown amount of nucleic acid of interest. The solution may be diluted to 10,000 ul and then divided into 1000 reaction volumes with each reaction volume (10 ul) containing about 0.1 ul of the original undiluted solution. If polynucleotide amplification reactions are performed in 1000 reaction volumes and about 90 reaction volumes are determined to have nucleic acid of interest amplified, the number of molecules present in the original undiluted solution is about 90 (dilution factor =1) and concentration of the nucleic acid of interest may also be determined if the molecular weight of the nucleic acid is known.

C. Other reactions

The instant method and apparatus may be used to investigate large numbers of molecular binding interactions, for example, sequence-specific interactions between polynucleotides (or modifications thereof) and polypeptides (or modifications thereof). The array assembly may be utilized to screen large numbers of polypeptides with different sequences for binding to a particular polynucleotide sequence or a large number of polynucleotides with different sequences for binding to a particular polypeptide. Either polynucleotides or polypeptides may be immobilized on one array, although it is possible that none is immobilized. Other molecular binding interactions, such as nucleic acid hybridization, inhibitor-enzyme, receptor-antagonist, drug-DNA, antibody-antigen, toxin-carbohydrate, protein-carbohydrate, and glycoprotein-receptor reactions may also be screened on a large scale using similar methods (see, e.g., Schullek et al., *Anal. Biochem.* 246:20–29 (1997), Pandey et al., *Nature* 405:837–846 (2000), de Wildt et al., *Nature Biotechnol.* 18:989 (2000), Uetz et al., *Nature* 403:623 (2000), and White *Annu. Rev. Pharmacol. Toxicol.* 40:133–157 (2000), all incorporated herein by reference).

For example, a combinatorial peptide library of potential enzyme inhibitors may be synthesized via an o-nitrobenzyl linker to a first solid support. Enzymes with buffers may be confined to reaction wells on a second solid support. After array assembly, potential enzyme inhibitors may be released by photolysis. A labeled enzyme substrate may be added to each reaction well. Enzyme inhibition may be identified and quantified using an optical system.

Polypeptides may be immobilized on solid supports and assayed for their functions. In one method, recombinant proteins (e.g., GST-fusion proteins) may be immobilized on a first solid support. Cell lysates containing interaction partners may be confined on a second array (e.g., in reaction wells). After array assembly, the system may be washed to remove unbound material. The bound proteins may be identified by known spectroscopic methods in the art. In another method, the yeast two-hybrid system may also be applied to array-based analysis. Yeast cells may be transformed with individual open reading frame-activation domain fusions and grown on a first array surface. Such array may then probed in a mating assay with yeast cells containing ORF-DNA-binding domain fusions on a second array.

IV. Probe Designs for Sequence Variation Detection or Polynucleotide Quantitation

In detecting sequence variations or quantitating polynucleotides, amplified nucleic acids are frequently determined by analyzing the extent of hybridization. The fundamental aspect of this method is the discrimination of hybridization stability between the match and the mismatch. A recurring problem to this discrimination is that a perfect match in A/T rich hybridization complexes would often have a lower stability than a mismatch in G/C rich hybridization complexes. This dependency of stability on base composition may lead to false positives or false negatives. Therefore, for successful and reliable detection, several factors in probe selection may be considered. See Cantor and Smith, *Genomics: the science and technology behind the human genome project*, John Wiley & Sons (1999).

A. Polynucleotide Sequence and Length

One of the factors influencing hybridization performance of a polynucleotide probe is base composition. It is well known that sequences rich in G/C are more stable than sequences with lower G/C content. The solution melting temperature (Tm) of a polynucleotide, at which 50% of the polynucleotide is hybridized and 50% is not hybridized, is often used as a practical indicator of the hybridization strength of a polynucleotide probe of a given base composition. Methods for measuring Tm of a polynucleotide are well known in the art. See, e.g., Cantor and Schimmel, *Biophycical Chemistry*, San Francisco, W. H. Freeman (1980), incorporated herein by reference. There are also many ways to calculate Tm using mathematical algorithm. A widely used rule of thumb is two degree of increase in Tm by adding an A/T base pair and four degree of increase in Tm by adding a G/C base pair. This simple formula may be further modified to take into account of the ionic strength and solvent effect. For example, Tm may be calculated using the formula:

$$Tm=81.5+16.6\ (\log Na^{+})+0.41\times\%\ \text{of G/C}-600/n-0.65\times\%\ \text{of formamide}$$

Where $Na^{+}$is sodium concentration, n is length of polynucleotide.

A more reliable formula to calculate Tm is available based on the interactions between a particular base and its nearest neighbors, i.e., the nearest-neighbor model. An enthalpy and entropy for each nearest neighbor combination of two adjacent base pairs (AA, AC, AG, AT, CA, CC, CG, CT, GA, GC, GG, GT, TA, TC, TG, and TT) have been established based on the extensive melting experiments using various polynucleotide sequences. Thermodynamic coefficients of nearest-neighbor models are available for DNA/DNA, DNA/RNA, and RNA/RNA hybridizations. Therefore, free energy of hybridization of two sequences at any temperature in solution may be calculated. See, e.g., U.S. Pat. No. 5,556,749, Hyndman, D., et al., *BioTechniques* 20(6) :1090–1097 (1996), Mitsuhashi, M., *J. Clinical Laboratory Analysis* 10:277–284 (1996), Wetmur, J., *Critical Reviews in Biochemistry and Molecular Biolog,* 26:227–259 (1991), Rychlik et al., *Nucleic Acids Res.* 17:8543–8551 (1989), and Rychlik et al., *Nucleic Acids Res.* 18:6409–6412 (1990), all incorporated herein by reference.

The hybridization behavior of immobilized polynucleotide probes on a solid support is different from that in solution. Therefore, a more empirical approach is necessary to predict and modulate hybridization behavior of array-immobilized polynucleotide probes. Additional melting temperature experiments on solid supports may be conducted to more accurately characterize the thermodynamics

and kinetics of hybridization behaviors of polynucleotide probes on an array. See Cantor and Smith, supra. Despite the differences in solid phase and solution phase kinetic and thermodynamic hybridization profiles, many variables affecting melting temperatures for solution hybridization, such as the effects of length, temperature, ionic strength, and solvent, are applicable for hybridization on solid supports.

Tm or free energy of hybridization may be evaluated based on base compositions, polynucleotide length, ionic strength, and thermodynamic parameters. High G/C content polynucleotide probes with a few mismatches may exhibit more stable hybridization than AT-rich polynucleotides without mismatches. Mismatches in the middle of the probe sequence are more consequential for hybridization than those at the 5' or 3' end. Shorter probe lengths may provide the maximum mismatch destabilization and result in the greater match to mismatch ratios. However, this advantage is partially offset by the wide range of Tm values for short probes, depending on their specific sequence composition. For example, probes with 17 nucleotides long with a single base difference may differ by 5° C. in Tm. If an array with equal length polynucleotide probes is used, baseline hybridization may yield wide range of signal intensities due to wide range of Tm values.

One skilled in the art will appreciate that in order to increase or decrease the melting temperature of a probe, it may be desirable to add, delete or change one or more bases in the probes. In certain embodiments of the inventions, polynucleotide probes with similar solution melting temperatures may be selected. The length of a polynucleotide probe may be changed, for example, by less than about 10, 5, 4, 3, 2 or 1 nucleotides.

Consideration of secondary structure may also play a role in evaluating hybridization performance of polynucleotide probes, especially when high hybridization temperature to denature secondary structures may not be applied. If polynucleotides form secondary structure such as hairpins or triple helixes, intramolecular hybridization within polynucleotides may be energetically and kinetically favorable and they may not be available for hybridization to the target sequences. See Mitsuhashi, M., *J. Clinical Laboratory Analysis, supra.*

In some instances, the presence of frequently appearing short subsequences may also be a factor for designing optimal polynucleotide sequences. For example, if polynucleotides contain a poly T or poly A stretch, such polynucleotides may cross-hybridize to poly(A)-mRNA or cDNA. If polynucleotides contain TATA-like sequences, such polynucleotides may bind to the promoter region of various genes.

A wide range of probe length may be used. Longer probes do not necessarily improve their sensitivity, because long probes usually exhibit higher Tm than that of actual assay conditions, allowing more mismatches. Although shorter probes increase the chances of nonspecific appearance of such sequences in the target sequences, they may exhibit a much higher penalty on mismatches. Therefore, one may design optimal probes based on their hybridization performance, instead of the length of the probes. In preferred embodiments of the present invention, the length of polynucleotide probes ranges from about 10 to about 100 nucleotides, preferably from about 10 to about 50 nucleotides, more preferably from about 15 to about 35 nucleotides.

B. Polynucleotide analogs

An alternative approach to even out base composition effects comprises the modification of one or more natural deoxynucleosides (or polynucleotide analogs) which forms a base pair whose stability is very close to that of the other pair. Polynucleotide analogs include base and sugar phosphate backbone analogs.

Any base analogs that induce a decrease in stability of the three G/C hydrogen bonds or an increase in stability of the two A/T hydrogen bonds may be used. For example, one can substitute 2,6-diamino purine for A, which gives 2-NH_2A/T base pair having a stability similar to that of the G/C base pair. One may also select C derivatives, in which one hydrogen of the exocyclic amino group at position 4 is substituted by an alkyl group such as methyl, ethyl, n-propyl, allyl or propargyl groups. For example, a $G^{4Et}C$ base pair has stability similar to that of the A/T base pair. Typically, it may be easier to find a modified G/C base pair whose stability is similar to that of an A/T natural base pair than to design a modified A/T base pair whose stability is close to that of a G/C natural base pair. In addition, preparation of polynucleotides containing C analogs may be simpler than that of polynucleotides built with G analogs and modification of only one base pair rather than both may simplify the preparation of polynucleotides containing one or several modified nucleosides. Analogs that increase base stacking energy, such as pyrimidines with a halogen at the C5-position (e.g. 5-bromoU, or 5-ChloroU), may also be used. One may also use the non-discriminatory base analogue, or universal base, such as 1-(2-deoxy-D-ribfuranosyl)3-nitropyrrole. This class of analogue maximizes stacking while minimizing hydrogen-bonding interactions without sterically disrupting a hybridization complex. See Nguyen, H., et al., *Nucleic Acids Research* 25(15) 3059–3065 (1997) and Nguyen, H., et al., *Nucleic Acids Research* 26(18):4249–4258 (1998), both incorporated herein by reference.

The highly charged phosphodiesters in natural nucleic acid backbone may be replaced by neutral sugar phosphate backbone analogues. The polynucleotide probes with uncharged backbones may be more stable, as in these analogs, the electrostatic repulsion between nucleic acid strands is minimized. As an example, phosphotriesters in which the oxygen that is normally charged in natural nucleic acids is esterified with an alkyl group may be used.

Another class of backbone analogs is polypeptide nucleic acids (PNAs), in which a peptide backbone is used to replace the phosphodiester backbone. The stability of PNA-DNA duplex is essentially salt independent. Thus low salt may be used in hybridization procedures to suppress the interference caused by stable secondary structures in the target. PNAs are capable of forming sequence-specific duplexes that mimic the properties of double-strand DNA except that the complexes are completely uncharged. Furthermore, because the hybridization stability of PNA-DNA is higher than that of DNA-DNA, binding is more specific and single-base mismatches are more readily detectable. See, e.g., Giesen, U. et al., *Nucleic Acids Research* 26(21):5004–5006 (1998), Good, L., et al., *Nature Biotechnology* 16:355–358 (1998), and Nielsen, P., *Current Opinion in Biotechnology* 10:71–75 (1999), all incorporated herein by reference.

Another option to modulate the hybridization performance of polynucleotide probes is the replacement of naturally occurring nucleic acids have 3'–5' phosphodiester linkage. Polyribonucleotides with 2'5' linkage which give complexes with lower melting temperature than duplexes formed by 3'–5' polynucleotides with the same sequence may be employed. See Kierzek, R., et al., *Nucleic Acids Research* 20(7):1685–1690 (1992), incorporated herein by reference.

Another method for optimizing hybridization performance is using polynucleotides containing C-7 propyne analogs of 7-deaza-2'-deoxyguanosine and 7-deaza-2'-deoxyadenosine (Buhr et al., *Nucleic Acids Res.* 24:2974–2980 (1996), incorporated herein by reference) or C-5 propyne pyrimidines (Wagner et al., *Science* 260:1510–3 (1993), incorporated herein by reference). These analogs may be particular useful in gene expression analysis.

C. Hybridization environment

Hybridization performance of polynucleotide is also dependent on the hybridization environment, for example, the concentrations of ions and nonaquous solvents. The hybridization performance of polynucleotide probes may be modulated by changing the dielectric constant and ionic strength of the hybridization environment. Salt concentrations, such as Na, Li, and Mg, may have an important influence on hybridization performance of polynucleotide probes.

Reagents that reduce the base composition dependence of hybridization performance may be used to alter the hybridization environment of array-immobilized polynucleotide probes. For example, high concentrations of tetramethylammonium salts (TMAC), N,N,N,-trimethylglycine (Betain) may be added to target nucleic acid mixture. At suitable concentrations typically at multimolar concentrations, these reagents may equalize the Tm of polynucleotides that are pure A/T and those that are pure G/C and thus increase the discrimination between perfect matches and mismatches. See, Von Hipppel et al., *Biochemistry*, 3:137–144 (1993), incorporated herein by reference.

Denaturing reagents that lower the melting temperature of double stranded nucleic acids by interfering with hydrogen bonding between bases may also be used.

Denaturing agents, which may be used in hybridization buffers at suitable concentrations (e.g. at multimolar concentrations), include formamide, formaldehyde, DMSO ("dimethylsulfoxide"), tetraethyl acetate, urea, GuSCN, and glycerol, among others.

Chaotropic salts that disrupt van der Waal's attractions between atoms in nucleic acid molecules may also be used. Chaotropic salts, which may be used in hybridization buffers at suitable concentrations (e.g. at multimolar concentrations), include, for example, sodium trifluoroacetate, sodium tricholoroacetate, sodium perchlorate, guanidine thiocyanate, and potassium thiocyanate, among others. See, Van Ness, J., et al., *Nucleic Acids Research* 19(19):5143–5151 (1991), incorporated herein by reference.

Renaturation accelerants that increase the speed of renaturation of nucleic acids may also be used. They generally have relatively unstructured polymeric domains that weakly associate with nucleic acid molecules. Accelerants include cationic detergents such as, CTAB ("cetyltrimethylammonium bromide") and DTAB ("dodecyl trimethylammonium bromide"), and, heterogenous nuclear ribonucleoprotein ("hnRP") A1, polylysine, spermine, spermidine, single stranded binding protein ("SSB"), phage T4 gene 32 protein and a mixture of ammonium acetate and ethanol, among others. See, Pontius, B., et al., *Proc. Natl. Acad. Sci. USA* 88:82373–8241 (1991), incorporated herein by reference.

One of skill in the art would appreciate that there are many other ways to modulate the hybridization performance of polynucleotides by changing the hybridization environment of polynucleotide probes. One method is changing the length of spacer that tethers polynucleotide probe to the array surface. It has been demonstrated that steric factors are important in increasing the efficiency of hybridization between polynucleotide probes and target nucleic acids. See, Southern et al., *Nucleic Acids Research*, 20(7):1679–1684 (1992), incorporated herein by reference. Methods for reducing non-specific binding to an array by surface modifications and probe modifications are described in WO 99/54509, incorporated herein by reference.

An alternative approach for enhancing the discrimination between matched and mismatches is applying electric current to polynucleotide probes which destabilize mismatches relative to matches. See, e.g., U.S. Pat. No. 5,929,208.

In some instances, the local concentration of polynucleotide probes or the concentration of target nucleic acids may be varied to allow maximum discrimination between matches and mismatches. In some instances, local concentrations of polynucleotide probes may be higher than target nucleic acids. Such high local DNA probe concentrations may generate high local charge densities and promote the undesirable association of probes that may interfere with target binding. High local probe concentration may also permit the simultaneous binding of target molecules to multiple probes, and may sterically prohibit access of target to the probes. If polynucleotide probes are at lower concentrations compared with the target sequence, the kinetics and thermodynamics of the hybridization may also be affected. See, Cantor and Smith, supra.

The polynucleotide probe set may also include control probes. One class of control probes is the mismatch probes. A mismatched probe is a probe whose sequence is deliberately selected not to be perfectly complementary to a reference sequence. In other words, mismatch probes are probes identical to their corresponding perfectly complementary probes except the presence of one or more mismatched bases. Therefore, under suitable hybridization conditions, the perfectly matched would be expected to hybridize with its target sequence, but mismatch probes would not hybridize or would hybridize to a significantly lesser extent, thus providing a control for non-specific binding or cross-hybridization. Although one or more mismatches may be located anywhere in the mismatch probe, probes are often designed to have the mismatch locate at or near the center of the probe such that the mismatch is most likely to destabilize the hybridization complex with the target sequence. In addition, the mismatch site is typically not the location of the sequence variation to be determined, but is within several nucleotides (e.g., less than 5) on the 5' or 3' side of the sequence variation location. For example, a probe set for a known biallelic SNP may contain two groups of mismatch probes based on two reference sequences constituting the respective polymorphic forms. Each group of mismatch probes may include at least two sets of probes, which each set contains a series of probes with a mismatch at one nucleotide 5' and 3' to the polymorphic site.

Control probes may also include normalization probes. Normalization probes are those perfectly complementary to a known polynucleotide sequence that is added to the target nucleic acids. Normalization probes provide a control for variation in hybridization condition, signal intensity, and other factors that may cause the signal of a perfect hybridization to vary between arrays. Normalization probes may be located throughout the array to control for spatial variation in hybridization intensity. For example, they may be located at the corners, edges or middle of the array.

The number of polynucleotide probes for a sequence variation or a gene expression may vary depending on the nature of sequence variation, gene expression, and level of

resolution desired. At least about 2, 5, 10, 20, or 50 polynucleotide probes may be employed for each sequence variation or each gene. Each probe in both sequence variation determination and gene profiling may be about 10 to 100 nucleotides long, e.g. shorter than about 20, 30, 40, 50, 60, 70, 80, or 90 nucleotides long. In the case of overlapping polynucleotides probes, the overlap may be about 1 to 50 bases, preferably below 30, 20, 10, or 5 bases.

V. Hybridization Conditions and Signal Processing

The hybridization can take place in the same reaction site or well as the amplification reactions. Generally, incubation may be at temperatures normally used for hybridization of nucleic acids, for example, between about 20° C. and about 75° C., e. g., above about 30° C. 40° C., 50° C., 60° C., or 70° C. The amplified nucleic acid may be incubated with the array for a time sufficient to allow the desired level of hybridization between the amplified product and any complementary probes in the array, usually in about 10 minutes to several hours. But it may be desirable to hybridize longer. After incubation with the hybridization mixture, the array is usually washed with the hybridization buffer, thus removing unhybridized molecules. This leaves only hybridized target molecules. Then the array may be examined to identify the polynucleotide probes to which the amplified product has hybridized.

Suitable hybridization conditions may be determined by optimization procedures or experimental studies. Such procedures and studies are routinely conducted by those skilled in the art. See e.g., Ausubel et al., *Current Protocols in Molecular Biology, Vol.* 1–2, John Wiley & Sons (1989) and Sambrook et al., Molecular Cloning A Laboratory Manual, 2nd Ed., Vols. 1–3, Cold Springs Harbor Press (1989). For example, hybridization and washing conditions may be selected to detect substantially perfect matches. They may also be selected to allow discrimination of perfect matches and one base pair mismatches. They may also be selected to permit the detection of large amounts of mismatches. As an example, the wash may be performed at the highest stringency that produces results and that provides a signal intensity greater than approximately 10% of the background intensity.

The hybridization intensities indicating the hybridization extent between the amplified nucleic acid and polynucleotide probes may be determined and compared. The differences in hybridization intensities are evaluated. One of skilled in the art will appreciate that methods for evaluating the hybridization results vary with the nature of probes, sequence variations, gene expressions, and labeling methods. For example, quantification of the fluorescence intensity is accomplished by measuring probe signal strength at locations where probes are present. Comparison of the absolute intensity of array-immobilized polynucleotide probes hybridized to amplified nucleic acids with intensities produced by control probes provides a measure of the sequence variations or the relative expression of the genes.

Quantification of the hybridization signal can be by any means known to one of skill in the art. For example, quantification may be achieved by the use of a confocal fluorescence microscope. The methods of measuring and analyzing hybridization intensities may be performed utilizing a computer. The computer program typically runs a software program that includes computer code for analyzing hybridization intensities measured. Signals may be evaluated by calculating the difference in hybridization signal intensity between each polynucleotide probe, its related probes, and control probes. The differences can be evaluated for each sequence variation or each gene.

Background signals typically contribute to the observed hybridization intensity. The background signal intensity refers to hybridization signals resulting from non-specific binding, or other interactions, e.g., between amplified nucleic acids and array surface. Background signals may also be produced by the array component itself. A single background signal may be calculated for an array or a different background signal may be calculated for each sequence variation or each gene expression analysis. For example, background may be calculated as the average hybridization signal intensity for the lowest 5% to 10% of the probes in the array, or where a different background signal is calculated for each sequence variation or gene, for the lowest 5% to 10% of the probes for each sequence variation or gene. Background signal may also be calculated as the average hybridization signal intensity produced by hybridization to probes that are not complementary to any sequence found in the sample (e.g. probes directed to nucleic acids of the opposite sense or to genes not found in the sample). Background may also be calculated as the average signal intensity produced by regions of the array that lack any probes at all. Preferably the difference in hybridization signal intensity between each probe and its control probes is detectable, e.g. greater than about 10%, 20%, or 50% of the background signal intensity. In some instances, only those probes where difference between the probe and its control probes exceeds a threshold hybridization intensity (e.g. preferably greater than 10%, 20%, or 50% of the background signal intensity) are selected. Thus, only probes that show a strong signal compared to their control probes are selected.

The identity of each sequence variation may be estimated using known methods in the art. If the target is present, the perfectly matched probes should have consistently higher hybridization intensity than the mismatched probes. Therefore, in sequence variation determinations, one of the four A, T, C, G substituted probes may have a significantly higher signal than the other three. A comparison of the intensities of four corresponding probes may reveal the identity of one sequence variation in the target sequence. For example, the highest intensity probe may be compared to the second highest intensity probe. The ratio of the intensities may be compared to a predetermined ratio cutoff, which is a number that specifies the ratio required to identify a sequence variation. For example, if the ratio cutoff is 1.2, a ratio of 1.4 is greater than the cutoff and the sequence variation may be determined. Of course, ratio cutoff may be adjusted to produce optimal results for a specific array and for a specific sequence variation.

In addition to comparing to mismatch probes, the hybridization intensity may be compared to other control probes, such as normalization probes. For example, probe intensity of amplified nucleic acid may be compared to that of a known sequence. Any significant changes may indicate the presence or absence of a sequence variation. Statistical method may also be used to analyze hybridization intensities in determining sequence variations or gene expression levels. For example, mismatch probe intensities may be averaged. Means and standard deviations may be calculated and used in determining sequence variations and profiling gene expressions. Complex data processing and comparative analysis may be found in EP 717 113 A2 and WO 97/10365, both incorporated herein by reference.

VI. Target Nucleic Acid Preparation and Reactant Labeling

It will be appreciated by one of skill in the art that the reactions using the instant invention may be monitored or quantitated by directly or indirectly labeling reactants, reac-

tion intermediates, or reaction products with detectable labels, prior to, during, or after the initiation of reactions. Detectable labels suitable for use in the present invention include any composition detectable by spectroscopic, photochemical, biochemical, immunochemical, electrical, optical or chemical means. Useful polynucleotide labels in the present invention include biotin for staining with labeled streptavidin conjugate, magnetic beads (e.g., Dynabeads™), fluorescent molecules (e.g., fluorescein, texas red, rhodamine, green fluorescent protein, FAM, JOE, TAMRA, ROX, HEX, TET, Cy3, C3.5, Cy5, Cy5.5, IRD41, BODIPY and the like), radiolabels (e.g., ^{3}H, ^{25J}I, ^{35}S, ^{34}S, ^{14}C, ^{32}P, or ^{33}P), enzymes (e.g., horse radish peroxidase, alkaline phosphatase and others commonly used in an ELISA), calorimetric labels such as colloidal gold or colored glass or plastic (e.g., polystyrene, polypropylene, latex, etc.) beads, mono and polyfunctional intercalator compounds. Means of detecting such labels are well known to those of skill in the art. For example, radiolabels may be detected using photographic film or scintillation counters, fluorescent markers may be detected using a photodetector to detect emitted light. Enzymatic labels are typically detected by providing the enzyme with a substrate and detecting the reaction product produced by the action of the enzyme on the substrate, and calorimetric labels are detected by simply visualizing the colored label.

Any nucleic acid specimen, in purified or non-purified form, can be utilized as the target sequence nucleic acid. The amplification reaction may amplify, for example, DNA or RNA, including mRNA, wherein DNA or RNA may be single stranded or double stranded. In the event that RNA is to be used as a template, enzymes and conditions optimal for reverse transcribing RNA to DNA may be utilized. In addition, a DNA-RNA hybrid which contains one strand of each may be utilized. A mixture of nucleic acids may also be employed, or the nucleic acids produced in a previous amplification reactions.

The target nucleic acids may be prepared from human, animal, viral, bacterial, fungal, or plant sources using known methods in the art. For example, target sample may be obtained from an individual being analyzed. For assay of genomic DNA, virtually any biological sample is suitable. For example, convenient tissue samples include whole blood, semen, saliva, tears, urine, fecal material, sweat, buccal, skin and hair. For assay of cDNA or mRNA, the tissue sample must be obtained from an organ in which the target nucleic acid is expressed. For example, if the target nucleic acid is a cytochrome P450, the liver is a suitable source. The target nucleic acids may also be obtained from other appropriate source, such as nuclear RNA, rRNA, tRNA, M 13, plasmid or lambdavectors, and cosmid or YAC inserts. The target may be preferably fragmented before application to the array to reduce or eliminate the formation of secondary structures in the target. The fragmentation may be performed using a number of methods, including enzymatic, chemical, thermal cleavage or degradation. For example, fragmentation may be accomplished by heat/Mg^{2+} treatment, endonuclease (e.g., DNAase 1) treatment, restriction enzyme digestion, shearing (e.g., by ultrasound) or NaOH treatment. Examples of target nucleic acid preparation are described in e.g., WO 97/10365, incorporated herein by reference.

Definitions

As used herein, the term "linker" refers to an anchoring group that serves to anchor or tether a molecule to a solid support during solid phase synthesis. The linker is sometimes the point of cleavage following synthesis.

As used herein, the terms "polynucleotide" and "nucleic acid" refer to naturally occurring polynucleotides, e.g. DNA or RNA. This term also refers to analogs of naturally occurring polynucleotides. The polynucleotide may be double stranded or single stranded. This term is used without referring to a specific length. The polynucleotides may be labeled with radiolabels, fluorescent labels, enzymatic labels, proteins, haptens, antibodies, sequence tags.

As used herein, the term "polynucleotide amplification reaction" refers to a broad range of techniques for increasing the number of copies of specific polynucleotide sequences. Typically, amplification of either or both strand of the target nucleic acid comprises the use of one or more nucleic acid-modifying enzymes, such as a DNA polymerase, a ligase, an RNA polymerase, or an RNA-dependent reverse transcriptase. Examples of polynucleotide amplification reaction include, but not limited to, polymerase chain reaction (PCR), nucleic acid sequence based amplification (NASB), self-sustained sequence replication (3SR), strand displacement activation (SDA), ligase chain reaction (LCR), Qβ replicase system, and the like.

As used herein, the term "polypeptide" refers to a polymer of amino acids without referring to a specific length. This term includes naturally occurring protein. The term also refers to modifications, analogues and functional mimetics thereof. For example, modifications of the polypeptide may include glycosylations, acetylations, phosphorylations, and the like. Analogues of polypeptide include unnatural amino acid, substituted linkage, etc. Polypeptides may be labeled with radiolabels, fluorescent labels, enzymatic labels, proteins, haptens, antibodies, sequence tags.

As used herein, the term "polynucleotide" refer to naturally occurring polynucleotide, e.g. DNA or RNA. This term does not refer to a specific length. Thus, this term includes oligonucleotide, primer, probe, gene, nucleic acid, etc. This term also refer to analogs of naturally occurring polynucleotides. Polynucleotides may be double stranded or single stranded. Polynucleotides may be labeled with radiolabels, fluorescent labels, enzymatic labels, proteins, haptens, antibodies, sequence tags, etc.

As used herein, the term "primer" refers to a polynucleotide, which is capable of annealing to a complementary template nucleic acid and serving as a point of initiation for template-directed nucleic acid synthesis, such as a polynucleotide amplification reaction. A primer need not reflect the exact sequence of the template but must be sufficiently complementary to hybridize with a template. Typically, a primer will include a free hydroxyl group at the 3' end. The appropriate length of a primer depends on the intended use of the primer but typically ranges from 12 to 40 nucleotides preferably from 15 to 40, most preferably from 20 to 40 nucleotides. The term primer pair (e.g., forward and reverse primers) usually means a set of primers including a 5' upstream primer that hybridizes with the 5' end of the target sequence to be amplified and a 3', downstream primer that hybridizes with the complement of the 3' end of the target sequence to be amplified.

As used herein, the term "probe" refers to a polynucleotide of any suitable length which allows specific hybridization to a polynucleotide. Probes may be attached to a label or reporter molecule. Typically, probes are at least about 10 nucleotides long.

As used herein, the term "reactant" refers to any component of a non-unimolecular reaction. A reactant may be a chemically or biologically reactive substance in a reaction.

As used herein, the term "sequence variation" of a polynucleotide encompasses all forms of polymorphism, muta-

tions and haplotypes. A sequence variation may range from a single nucleotide variation to the insertion, modification, or deletion of more than one nucleotide. A sequence variation may be located at the exon, intron, or regulatory region of a gene.

Polymorphism refers to the occurrence of two or more genetically determined alternative sequences or alleles in a population. A biallelic polymorphism has two forms. A triallelic polymorphism has three forms. A polymorphic site is the locus at which sequence divergence occurs. Diploid organisms may be homozygous or heterozygous for allelic forms. Polymorphic sites have at least two alleles, each occurring at frequency of greater than 1% of a selected population. Polymorphic sites also include restriction fragment length polymorphisms, variable number of tandem repeats (VNTR's), hypervariable regions, minisatellites, dinucleotide repeats, trinucleotide repeats, tetranucleotide repeats, simple sequence repeats, and insertion elements. The allelic form occurring most frequently in a selected population is sometimes referred to as the wild type form or the consensus sequence.

Mutations include deletions, insertions and point mutations in the coding and noncoding regions. Deletions may be of the entire gene or of only a portion of the gene. Point mutations may result in stop codons, frameshift mutations or amino acid substitutions. Somatic mutations are those which occur only in certain tissues, such as liver, heart, etc and are not inherited in the germline. Germline mutations can be found in any cell of a body and are inherited.

A haplotype refers to the combination of sequence variations that co-exist on a chromosome.

The term "subsequence" refers to a polynucleotide sequence that comprises a part of a longer polynucleotide sequence.

Examples of the Preferred Embodiments

The following examples further illustrate the present invention. These examples are intended merely to be illustrative of the present invention and are not to be construed as being limiting.

EXAMPLE 1

Synthesis of DMT-protected o-nitrobenzyl Amidite

A schematic illustration of DMT-protected o-nitrobenzyl amidite synthesis is shown in FIG. 7.

6-(4-Formyl-2-methoxyphenoxy)hexan-1-ol (2)

A flame-dried, 100 mL round bottom flask equipped with a claisen arm and reflux condenser was charged, under argon, with vanillin (1, 1.0 g, 6.6 mmol), tetrabutylammonium iodide (369 mg, 1 mmol), and potassium carbonate (1.38 g, 10 mmol). These were suspended in 10 mL of anhydrous acetonitrile and then 6-chlorohexanol (1.05 ml, 7.9 mmol) was added dropwise. The reaction was protected from light and refluxed for 15 hours at which time TLC analysis showed complete consumption of the starting material. The solvent was evaporated in vacuo and the residue was then dissolved in 50 mL of ethyl acetate and extracted with water (2×100 mL), brine (1×100 mL), and water (1×100 mL). The organic layer was dried over magnesium sulfate, filtered, and the solvent was evaporated in vacuo to give 1.5 g (90%) of 2 as a white solid. R_f=0.29 (33% ethyl acetate/hexane). Mp 55.0–56.0° C. ^{1}H NMR (400 MHz, DMSO-d6): δ9.83 (s, 1), 7.53 (dd, J=2.0 Hz, 8.0 Hz, 1), 7.39 (d, J=2.0 Hz, 1), 7.17 (d, J=8.4 Hz, 1), 4.36 (t, J=5.2 Hz, 1), 4.07 (pt, J=2.4 Hz, 2.8 Hz, 2), 3.83 (s, 3), 3.39 (dd, J=5.6 Hz, 11.6 Hz, 2), 1.75 (m, 2), 1.45–1.34 (m, 6) ppm.

6-(4-Formyl-2-methoxyphenoxy)hexyl acetate (3).

A flame-dried, 250 mL round bottom flask was charged, under argon, with 2 (10 g, 39.6 mmol) and this was dissolved in 50 mL of anhydrous pyridine and the flask was protected from light. To this solution was added acetic anhydride (9.3 mL, 98.8 mmol) dropwise at a rate of 0.5 mL/minute. The solution was stirred at room temperature for 30 minutes at which time all of the starting material had been consumed. The solvent was evaporated with reduced pressure and then toluene was added to remove excess pyridine. Evaporation of the solvent in vacuo gave 11.5 g (99%) of 3 as an off-white solid. R_f=0.74 (33% ethyl acetate/hexane). Mp 74.0–75.0° C. ^{1}H NMR (400 MHz, DMSO-d6): δ9.83 (s, 1), 7.53 (dd, J=1.6 Hz, 8.4 Hz, 1), 7.39 (d, J=2.0 Hz, 1),), 7.17 (d, J=8.4 Hz, 1), 4.07 (t, J=6.4 Hz, 2), 4.00 (pt, J=6.4 Hz, 6.8 Hz, 2), 3.83 (s, 3), 2.00 (s, 3) 1.74 (m, 2), 1.58 (m, 2), 1.4–1.35 (m, 4) ppm.

6-(4-Formyl-2-methoxy-5-nitrophenoxy)hexyl acetate (4).

A 250 mL round bottom flask was charged with 3 (10 g, 34 mmol) and was then placed in an ice bath. To this was added a 20% v/v solution of fuming nitric acid in acetic acid (40 mL total volume) at a rate of 1.0 mL/minute with stirring. At the end of addition the bath was removed, the reaction was protected from light, and stirred at room temperature for 4 hours. At this time the solution was poured over crushed ice (200 mg). Water was added and the product was extracted with ethyl acetate (3×200 mL). The combined organics were washed with brine (1×200 mL), saturated sodium bicarbonate (1×200 mL), and water (1×200 mL), dried over magnesium sulfate, and filtered. The solvent was evaporated in vacuo to give 10.6 g of 4 (92%) as an orange-red solid. R_f=0.80 (33% ethyl acetate/hexane). Mp 79.8–81.5° C. ^{1}H NMR (400 MHz, DMSO-d_6): δ7.69 (s, 1), 7.37 (s,1), 4.17 (pt, J=6.4, 6.8 Hz, 2), 4.0 (t, J=6.8 Hz, 2), 3.95 (s,3), 1.99 (s, 3), 1.77 (m, 2), 1.59 (m, 2), 1.45–1.37 (m, 4) ppm.

6-(4-Hydroxymethyl-2-methoxy-5-nitrophenoxy)hexyl acetate (5)

A flame-dried, 250 mL round bottom flask was charged, under argon, with 4 (8 g, 23.6 mmol) and this was suspended in 35 mL of anhydrous ethanol. The flask was protected from light and then placed in an ice bath. Sodium borohydride (2.06 g, 54.3 mmol) was added in small portions and the reaction was allowed to stir at 0° C. for 180 minutes. At this time, saturated ammonium chloride (30 mL) was added slowly to the reaction mixture over a period of 20 minutes to minimize the gas evolution. Ethanol was evaporated in vacuo and the mixture was diluted with water (100 ml) and extracted with dichloromethane (3×100 mL). The combined organic layers were dried with magnesium sulfate, filtered, and the solvent was evaporated in vacuo to give orange oil. This oil was subjected to column chromatography (SiO_2, 75×200 mm, 1% triethylamine in dichloromethane) to give, following solvent evaporation, 7.25 g (90%) of 5 as a yellow solid. R_f=0.30 (33% ethyl acetate/hexane). Mp >220° C. ^{1}H NMR (400 MHz, DMSO-d_6): δ7.67 (s, 1), 7.38 (s, 1), 5.55 (t, J=6.4 Hz, 1), 4.81 (d, J=6.4 Hz, 2), 4.04 (t, J=7.2 Hz, 2), 4.00 (t, J=7.6 Hz, 2), 3.91 (s, 3), 1.99 (s, 3), 1.76–1.72 (m, 2), 1.61–1.57 (m, 2), 1.45–1.37 (m, 4) ppm.

6-[4-(4,4'-Dimethoxytrityloxymethyl)-2-methoxy-5-nitrophenoxy)hexan-1-ol (7).

A flame-dried, 250 mL round bottom flask was charged, under argon, with 5 (5 g, 14.7 mmol) and this was dissolved in 45 mL of anhydrous pyridine. The flask was protected from light and 4,4'-dimethoxytrityl chloride (7.4 g, 22 mmol) was added in one portion. The solution was stirred at room temperature for 18 hours. Toluene was then added to the reaction flask and the solvent was evaporated in vacuo to

give an oil. This process of toluene addition/evaporation was repeated twice and then water was added to the flask to remove the last trace of organic solvents. The residue was dissolved in 400 mL of ethyl acetate and water extractions (3×150 mL) were performed. The organic layer was dried over magnesium sulfate, filtered, and the solvent was evaporated in vacuo to give 6 as a yellow oil. This was taken to the next step without further manipulation. R_f=0.68 (33% ethyl acetate/hexane). ^{1}H NMR (400 MHz, DMSO-d_6): δ7.60 (s, 1), 7.45–7.24 (m, 10), 6.90 (d, J=8.4 Hz, 4), 4.50 (s, 2), 4.04 (t, J=7.2 Hz, 2), 4.00 (t, J=7.6 Hz, 2), 3.95 (s, 3), 3.73 (s, 6), 1.99 (s, 3), 1.76–1.72 (m, 2), 1.61–1.57 (m, 2), 1.45–1.37 (m, 4) ppm.

A flame-dried, 250 ml round bottom flask was charged with 6 and this was dissolved in 100 mL of anhydrous methanol. The flask was protected from light and then potassium carbonate (2.6 g, 18.6 mmol) was added in one portion. The reaction was continued at room temperature for 30 minutes at which time TLC analysis showed complete consumption of the starting material. The solvent was evaporated in vacuo and the residue was then dissolved in 600 mL of dichloromethane and extractions performed with water (3×400 mL). The organic layer was dried over magnesium sulfate, filtered, and the solvent evaporated in vacuo to give a yellow oil. This oil was subjected to column chromatography (SiO_2, 75×200 mm, 1% triethylamine in dichloromethane) to yield, following solvent evaporation, 6.6 g (75%, 2 steps) of 7 as a waxy yellow solid. R_f=0.40 (33% ethyl acetate/hexane). ^{1}H NMR (400 MHz, DMSO-d_6): δ7.60 (s, 1), 7.45–7.24 (m, 10), 6.90 (d, J=8.4 Hz, 4), 4.35 (t, J=4.8, 5.2, 2), 4.03 (pt, J=6.4, 6.8 Hz, 2), 3.95 (s, 3), 3.73 (s, 6), 3.73–3.71 (m, 2), 1.95–1.98 (m, 2), 1.44–1.36 (m, 6) ppm. ^{13}C NMR (100 MHz, DMSO-d_6): δ158.6, 153.9, 146.9, 145.1, 139.1, 135.8, 130.3–127.3 (m), 113.8, 113.2, 110.4, 109.3, 87.0, 69.1, 63.0, 61.0, 56.4, 55.5, 32.9, 28.9, 25.7, 25.6 ppm.

6-[4-(4,4'-Dimethoxytrityloxymethyl)-2-methoxy-5-nitrophenoxy) hexyloxy-1-O-[O-(2-cyanoethyl)-N,N-diisopropylphosphoramidite (8).

A flame-dried, 250 ml round bottom flask was charged with 7 (6 g, 11.6 mmol). The system was protected from light and evacuated under reduced pressure for at least six hours. The system was back-filled with argon and 7 was dissolved in 50 mL of anhydrous dichloromethane. The system was placed in an ice bath and then diisopropylethylamine (10 mL, 58 mmol) was added followed by slow addition of 2-cyanoethyl-N,N-diisopropylchlorophosphoramidite (3.8 mL, 17.4 mmol). The ice bath was then removed and the reaction stirred at room temperature for 60 minutes. At this time, the reaction was quenched via the addition of 30 mL of methanol. The solvents were evaporated in vacuo and the residue was dissolved in dichloromethane and extracted with saturated sodium bicarbonate (1×300 mL). The organic layer was dried over magnesium sulfate, filtered, and the solvent evaporated in vacuo to give an orange-yellow oil. This oil was subjected to column chromatography (SiO_2, 75×200 mm, 1% triethylamine in dichloromethane) to yield, following solvent evaporation, 3.2 g (40%) of 8 as a faint orange oil. R_f=0.70 (33% ethyl acetate/hexane). ^{1}H NMR (400 MHz, DMSO-d_6): δ7.60 (s, 1), 7.42–7.46 (m, 3), 7.21–7.37 (m, 7), 6.9 (d, J=8.4 Hz, 4), 4.50 (s, 2), 4.03 (t, J =6.4 Hz, 2), 3.95 (s, 3), 3.73 (s, 6), 3.71–3.73 (m, 2), 3.52–3.6 (m, 6), 2.74 (t, J=6.0 Hz, 2), 2.44 (m, 2), 1.75–1.72 (m, 2), 1.58–1.55 (m, 2), 1.41–1.39 (m, 2), 1.18–1.10 (m, 12) ppm. ^{31}P NMR (162 MHz, DMSO-d_6): δ148.9, 147.1 ppm. ^{13}C NMR (100 MHz, DMSO-d_6): δ158.8, 154.1, 147.1, 145.2, 139.3, 137.9, 136.0, 130.4, 130.1, 129.4, 128.7, 128.6, 128.1, 127.4, 125.8, 119.5, 113.9, 113.3, 110.6, 109.5, 87.1, 69.2, 63.6, 63.4, 63.1, 58.9, 58.7, 58.6, 56.6, 55.6, 55.4, 50.9, 50.7, 43.1, 42.9, 31.2, 31.1, 29.0, 25.6, 25.5, 25.0, 24.9, 24.8, 21.6, 20.4 ppm.

EXAMPLE 2

Polynucleotide Synthesis off of a Photocleavable Linker

Aminoalkylated controlled-pore glass resin was treated with a 20% v/v solution of piperidine in anhydrous dimethylformamide (DMF) at room temperature for 30 minutes. Next, the resin was washed in anhydrous DMF and a 0.1 M solution of N-α-(9-fluorenylmethoxycarbonyl)-O-trityl-L-homoserine (Fmoc-HoSer(Trt)-OH) in anhydrous DMF containing 0.1 M N-[(dimethylamino)-1H-1,2,3-triazol[4,5-b]pyridin-1-ylmethylene]-N-methylmethanaminium hexafluorophosphate N-oxide (HATU) and 0.2 M anhydrous diisopropylethylamine (DIEA) was introduced and allowed to react for 30 minutes at room temperature. The resin was treated with a 20% v/v solution of piperidine in anhydrous DMF for 15 minutes to remove the Fmoc group. The resin was washed in anhydrous DMF and then treated with a 25% v/v solution of acetic anhydride in pyridine for 30 minutes. The resin was then treated with a 3% w/v solution of trichloroacetic acid (TCA) in dichloromethane (DCM) for 10 minutes to remove the trityl group and give a terminal hydroxyl group to couple to compound 8.

Coupling of 8 to the homoserine derivatized resin was accomplished through standard amidite chemistry (5-ethylthiotetrazole was used as the activator). Following benzylic DMT removal, a mixed sequence of 3'-CTC AAC CCG AAT CTC CG-5' (SEQ. ID. No. 1) was synthesized in the standard fashion on a 200 nmol scale. At the completion of the synthesis, the resin was treated with an aqueous solution of ammonia in order to deprotect the phosphate backbone and remove the heterocyclic protecting groups. No polynucleotide was detected in the eluent of these deprotection solutions. A portion of the resin was suspended in 18 MΩ water and subjected to photolysis. Using a converted lithography system, we delivered 16 mW/cm^2 (measured at 365 nm) of light for approximately 20 minutes. Cleavage of this polynucleotide from the resin gives a 3'-phosphate-terminated sample. It should be noted that no special purification was attempted (i.e. desalting column, HPLC, etc). An aliquot of the aqueous solution was then subjected to capillary electrophoresis. Capillary electrophoresis was carried out on a Beckman P/ACE 5000 system. Samples were injected electrokinetically onto a 37 cm×75 μm i.d. J&W Scientific 3% T 3% C μPAGE polyacrylamide gel filled column. Separation was done at 9 kV and the separation buffer used was Tris-borate pH 8.0 with 1M urea. The result is shown in FIG. 8. In addition, the sample was mixed with an authentic sample purchased from IDT, Inc (Coralville, IA, standard desalting purification) and subjected to CE analysis (FIG. 9). As can be seen in FIG. 9, the polynucleotides prepared from the resin terminated in 8 is as pure as that obtained commercially.

EXAMPLE 3

Preparation of Surface Derivatized Arrays

Glass slides were cleaned by sonication in a 2% solution of Micro 90 in Mill-Q water for 60 minutes at room temperature. The slides were then rinsed excessively with Milli-Q water and dried down with nitrogen. Next, the slides were exposed to an RF oxygen plasma (Plasmline421, Tegal, Novato, Calif.) for 60 minutes at 150 watts, 0.4 Torr, and 3.5 cc/min flow rate (Brzoska et al., *Nature* 360:719–721 (1992). The slides were further washed for 10' in a perox-

ysulfuric acid solution (70% H_2SO_4:15%H_2O_2, VWR, San Francisco, Calif.).

Immediately after oxygen plasma treatment, the slides were silanated with a 0.4% solution of 4-aminobutyldimethylmethoxysilane (ABS) (United Chemical Technologies, Bristol, Pa.) in anhydrous toluene (Aldrich, Milwaukee, Wis.) in a glove box under argon for 72 hours. The slides were then washed in anhydrous toluene with sonication for 15 minutes then rinsed in 95% ethanol (Aldrich) with sonication for 15 minutes. After drying each slide under nitrogen, the slides were cured for 30 minutes at 120° C. in an air oven.

Next, slides were coated with 3.5 micron layer of a positive photoresist (Microposit 1818; Shipley, Marlborough, Mass.) by spin coating photoresist (3.4 mls) at 1250 rpm for 30 seconds. After spin coating, the photoresist was soft baked for 30 minutes at 90° C. in an air oven. Next, the slides were photomasked by placing each slide onto a chromium mask (Image Technology, Palo Alto, Calif.) that had round features with each feature being 1 mm in diameter with center to center spacing of 2.0 mm with the photoresist side touching the mask. The chips were then exposed to near UV irradiation with a 365 nm 500 W columnated mercury lamp (45 mW/cm^2, AB-M, San Jose, Calif.) for 1.0 second. After exposure, the exposed photoresist was removed by placing the slides into a solution of Microposit 350 developer (1:1 in H_2O, Shipley, Marlborough, Mass.) for 30 seconds with agitation and then rinsed extensively with Milli-Q water and dried under argon.

The slides were then exposed to an RF plasma (Plasmline421) 150 watts, 0.4 Torr, and 3.5 cc/min flow rate, for 6.0 minutes to remove the ABS along with residual photoresist from the photolyzed regions. Next, the slides were silated with a 0.25% solution of (tridecafluoro-1,1,2,2-tetrahydrooctyl)-1-trichlorosilane (United Chemical Technologies, Bristol, Pa.) in anhydrous toluene (Aldrich) in an argon dry box for 10 minutes at room temperature and then washed in anhydrous toluene with sonication for 15 minutes. Finally, the photoresist covering the synthesis regions was stripped by sonication in acetone (Aldrich) then by washing in NMP (Aldrich) for 60 minutes at 70° C. and washed extensively in Milli-Q H_2O to remove residual photoresist.

EXAMPLE 4

In situ Synthesis of Polynucleotides with a Cleavable Site

All derivatizations were performed in a chip reaction chamber. Briefly, two slides were placed with their patterned surfaces facing one another. The gasket and chips were secured by four steel binder clips. Reagents were introduced via syringe through a 27-gauge needle. The gas interior was displaced through an open 27-gauge needle while reagent was being injected with the syringe. At the end of the given reaction, the reagent was removed via syringe using an open 27-gauge needle for venting. For this process, all washings between steps were done by first disassembling the reaction chamber and then rinsing each slide individually with the given solvent. Excess solvent was removed from the surface by means of a nitrogen gas stream. A fresh gasket was used for each subsequent chamber assembly and derivatization process.

Derivatized glass surfaces were patterned as follows: DMT-hexa-ethyloxy-glycol-CED phosphoramidite (DMT-HEG-CEP) was coupled to surface bound amines using a 1:1 solution of 0.1M linker and 0.45 M 5-ethylthiotetrazole in acetonitrile for 15 minutes with mixing. After two acetonitrile washes the chips were treated with a 0.1 M solution of iodine in tetrahydrofuran (THF)/pyridine/water for 1 minute and then washed twice with acetonitrile. Any uncoupled amines were acylated by treatment with a 25% v/v solution of acetic anhydride in pyridine for 15 minutes. Following DMT removal via treatment of the surfaces with a 3% w/v solution of trichloroacetic acid (TCA) in dichloromethane (DCM), compound 8 was coupled in some of the derivatized sites to the surface-bound terminal hydroxyls using standard phosphoramidite chemistry. In other derivatized sites, dT-CE phosphoramidite was coupled to serve as non-photolyzable controls. In every hydrophilic site, the following polynucleotide was prepared using standard phosphoramidite chemistry: 3'-GCA TGC ATG CAT GCA-5' (SEQ. ID. No. 2). Then, Cy3-phosphoramidite (Glen Research, Sterling, Va.) was coupled to every other derivatized site to provide a detectable, fluorescent end-label. The substrate-bound polynucleotides were then deprotected using a 1:1 v/v solution of ethylenediamine in ethanol.

After dicing the substrate into 1"×2" chips, a chamber assembly was made for the photolysis studies. Briefly, two chips (one derivatized, one blank) were placed with the patterned surface facing inwards and separated by a silicone rubber gasket. The gasket and chips were secured by four steel binder clips. Buffer (26 mM Tris-HCl, pH 9.5, 6.5 mM $MgCl_2$) was introduced via syringe through a 27-gauge needle. The gas interior was displaced through an open needle while reagent was being injected with the syringe. Photolysis was performed using a high-pressure mercury lamp with a light intensity of 16.0 mW/cm^2 centered at 365 nm. At the end of the given process, the buffer was removed via syringe. The chamber was disassembled and the chip was washed with acetonitrile, ethanol, then water and the fluorescence detected using an Axon GenePix 4000 scanner. As shown in FIG. **10**, the fluorescent signal in the photolabile sites decreases over time relative to the non-photolyzable sites. At or about 10 minutes, the signal decrease levels out and does not significantly lower over the next 120 minutes. This experiment shows that compound 8 couples to the surface through a linker and is a viable substrate for subsequent polynucleotide synthesis.

EXAMPLE 5

Photolytic Release of Polynucleotides from the Surface and Subsequent Hybridization of the Released Polynucleotides

This experiment was designed to address the issue of photolytic release of in situ synthesized polynucleotides from the array surfaces and study their subsequent hybridization to complimentary, surface-bound polynucleotides. This study was designed to address the issue of photolysis and subsequent motility and hybridization of the released polynucleotides from a linker derivatized with amidite 8.

A patterned, surface-tension array was prepared with 5×5 derivatized sites as depicted below (25 derivatized sites per unit cell, 10 total unit cells/chip). In the sites containing the "F" sequence, amidite 8 had been coupled. All other sites had a noncleavable amidite (dT-CEP, Glen Research, Sterling, Va.) coupled to the site. The synthesis was performed in a 3'-to-5' directions using commercially available reagents and house made robotics. The "F" polynucleotides were then end-labeled with Cy 3-Phosphoramidite (Glen Research, Sterling, Va.). Following deprotection of the polynucleotide backbone and heterocycles, the chip was scanned on an Axon GenePix4000. After the initial scan, the arrayed chips were assembled in a sandwich-like format with a microfabricated coverslip containing 10 wells (5×5×0.1 mm) that line up over the 10 unit cells of in situ synthesized polynucleotides. Each microfabricated well contained, nominally, 2.5 μL of 260 mM Tris-HCl, pH 9.5, 65 mM $MgCl_2$, 20 mM KCl, 0.1% Triton X-100. The coverslip and

polynucleotide immobilized chip were conjoined by use of 100 μL of a synthetic polymer (2×Self-Seal, MJ Research, Watertown, Mass.) deposited on the perimeter of the assembly. At this time, the assembly was exposed to ~200 J of 365 nm light. Following chamber disassembly and washing, the chip was scanned again and the signal intensities recorded and charted in FIG. 11. As can be seen, polynucleotides were released from the intended sites and did form a duplex with its compliment in the designated sites ("R" polynucleotides). The signal in the noncleavable sites increased 10 fold over the pre-photolysis numbers. The "C-" blank control was not hybridized to, as expected. It should be noted that the absolute fluorescence values were normalized to give a comparative representation.

This experiment was also performed under the following conditions. Using the same substrates, the array surface was exposed to 365 nm light (measuring at 16.0 mW/cm^2) for 15 minutes. Then, 30 μL of SSC buffer, pH 8.2, 50 mM sodium bicarbonate, 0.8% Tween-20, was floated over the surface with the use of a microscope cover slip (Fisher Scientific). After 30 minutes, the coverslip was removed, the array washed with briefly with 150 mM aqueous sodium bicarbonate solution, dried under a stream of argon, and scanned on an Axon GenePix4000. The results mirrored those described above with the exception that the hybridization signal increased 30 fold over the pre-photolysis numbers.

A 5 x 5 unit cell

F	F	F	F	F
F	R	F	R	F
F	F	C-	F	F
F	R	F	R	F
F	F	F	F	F

Abbreviations

Probe	3'-5' Sequence (SEQ. ID. Nos: 3–5)
F	TTTTATCGGAGATTCGGGTTGAG
R	ctcaacccgaatctccgataaaa
C-	GATGCTACCGTGACTGACTGACTGACTGA

EXAMPLE 6

In situ Synthesis of Probes for Sequence Variation or Gene Expression Detections

Drop-on-demand polynucleotide synthesis was performed on a DNA microarray synthesizer using the following -reagents (all reagents were purchased from Glen Research, Sterling, Va., unless noted): phosphoramidites: pac-dA-CE phosphoramidite, Ac-dC-CE phosphoramidite, iPr-pac-dG-CE phosphoramidite, dT CE phosphoramidite (0.1M); activator: 5-ethylthio tetrazole (0.45M). Amidites and activator solutions were premixed, 1:1 :v/v, in a 90% adiponitrile (Aldrich): 10% acetonitrile solution prior to synthesis. The following ancillary reagents were used: Oxidizer (0.1M iodine in THF/pyridine/water), Cap mix A (THF/2,6-lutidine/acetic anhydride), Cap mix B (10% 1-methylimidazole/THF), and 3% TCA in DCM. Parallel synthesis of individual polynucleotides was achieved by the addition of individual amidites to the hydrophillic regions of prepared surface tension arrays via custom designed piezo electric ink-jet devices (Microfab Technologies, Plano, Tex.). The jets were run at 6.67 kHz using a two step wave form which fired individual droplets of approximately 50 picoliters per drop.

For the 1 mm diameter features approximately 400 drops were added to each feature per nucleotide addition. After a suitable coupling time, the uncoupled amidites were washed off of the surface by flooding with acetonitrile then removed by spinning the chip at 2000 rpm for several seconds. All other reagents were added to the surface by flooding the substrate and removed after suitable reaction times. The synthesis was done in a closed nitrogen saturated environment with a unidirectional flow of protecting gas. The synthesis cycle is summarized in Table 1.

TABLE 1

Synthesis cycle for the production of a surface tension polynucleotide microarray.

Step in Cycle	Volume (mls)	Time (seconds)
ACN Wash	4.2	5
[1]Spin		5
ACN Wash	4.2	5
Spin		5
Deblock (3% TCA in DCM)	3.5	15
Spin		5
Deblock (3% TCA in DCM)	3.5	15
Spin		5
ACN Wash	4.2	5
Spin		5
[2]Dry down		10
Couple Amidites	2×10^{-6}	120
ACN Wash	4.2	5
Spin		5
Cap (CapA:CapB, 1:1)	4.5	15
Spin		5
Oxidize	5.5	15
Spin		5
Cap (CapA:CapB, 1:1)	4.5	10
Spin		5
Repeat cycle until the desired probes are produced		

[1]Spin speed between successive washing steps was 2000 rpm. Washing and coupling steps were done at different locations on the chip synthesizer where there was a continuous unidirectional flow of nitrogen that was directed from the synthesis location towards the washing position. This kept the ancillary reagent vapors from interacting with the amidites.
[2]A dry down was a step that was included prior to coupling to evaporate any residual ACN that may have been left on the hydrophilic regions of the array. This consisted of a high pressure nitrogen purge over the surface of the substrate.

EXAMPLE 7

Microfabrication of Wells on Arrays (coverslips)

Borofloat glass rounds (100 mm diameter and 0.7 mm thick) were used to prepare the coverslips. Several coverslip designs were made according to FIG. 12 and according to the following procedure:

Glass Cleaning and Patterning

A pre-cleaning step was performed by soaking the wafers in peroxysulfuric acid followed by a rinse and drying. The wafers were dipped in 49% hydrofluoric acid for about 15 seconds to roughen and clean the surface followed by rinsing and drying. The wafers were coated with a thin layer of amorphous silicon by chemical vapor deposition in a Tylan type oven at 525° C. at a thickness of 1500Å. To remove any moisture, the wafers were singe baked in an oven for about 30 minutes at 150° C. The wafers were primed with a layer of HMDS (hexamethyldisilazane) at 6000 RPM for 20 seconds before photo resist application. A viscous photo resist (Shipley SPR220-7) was used to coat the wafers and are spun at 3500 RPM for 30 seconds to a thickness of 7 um. The photo resist layer was pre-baked for about 200 seconds on a hot plate at 90° C. The wafers were then allowed to stand for about 3 hours prior to exposure. A UV aligner with a wavelength set at 350 nm with an exposing time of about 12 seconds was used for each wafer.

Due to the thickness of the photo resist, the wafers were allowed to stand for about 3 hours before developing. This helps improve critical line resolution of the exposed areas. The wafers were then dipped in developer solution (Shipley LDD26W) at least 3 times for about two minutes each time with a deionized water rinse after developing. A post-bake step wasperformed at 90° C. for one hour in an oven.

Glass Etching

The etching steps involve the pattern etching by removing the amorphous silicon layer and the actual glass etching. The pattern etching was accomplished by using a plasma etcher that uses sulfur hexafluoride and Freon 115 at a rate of 5.2 and 0.5 cm^3/s respectively. An etching time of 2 minutes was sufficient to remove the amorphous silicon. The wafers were etched in 49% hydrofluoric acid for an initial time of 8 minutes. A water rinse and spin dry step followed and the wafers were placed back into the 49% hydrofluoric acid solution for the remaining time required to reach a final depth of 100 microns. The wafers were rinsed and dried. The photo resist was removed using acetone and the remaining amorphous silicon was removed with heated potassium hydroxide at about 70° C. The wafers were then rinsed and dried. To remove the potassium on the glass surface, the wafers were treated with about a 20% solution of hydrochloric acid. A final peroxysulfuric acid treatment was performed to remove any residuals. The wafers were diced to a 1"×2" sized coverslips.

Chemical Derivatization of the Coverslips

A multiple cleaning process was performed prior to the aminosilation and acetylation of the coverslips. The coverslips were sonicated for about 10 minutes and water rinsed to remove any dust or glass particulates from the dicing process. To remove any dirt or oils on the glass surfaces, the coverslips were soaked with 10% weight to volume sodium hydroxide at 70° C. for 30 minutes. The coverslips were water rinsed and soaked in peroxysulfuric acid treatment for about an hour. To free the Si-O groups on the glass surface, the coverslips were placed in a plasma etcher with oxygen as the gas for 15 minutes at 70° C. The coverslips were placed in a reaction kettle and 1 ml of 3-aminopropyldimethylethoxy silane was added. The kettle was placed in an oven at 55° C. and left overnight to complete the gas phase reaction. The coverslips were removed from the kettle and were placed in an oven at 90° C. for an hour to cure the surface. The coverslips were sonicated in a 1:1 mixture of acetonitrile and ethanol for about 15 minutes to remove any un-reacted reagent.

To complete the acetylation, the coverslips were reacted with 25% (by volume) acetic anhydride with pyridine as the solvent. The coverslips were allowed to stand for at least two hours for the reaction to complete. The coverslips were sonicated with a 1:1 solution of acetonitrile and ethanol to remove any un-reacted acetic anhydride for about 15 minutes. The coverslips were dried with nitrogen and were ready for use in the array assembly.

EXAMPLE 8

Polymorohisms, Alleles, and Phenotypes of the NAT2 Gene

N-acetyltransferase 2 (NAT2) is a polymorphic N-acetylation enzyme that detoxifies hydrazine and arylamine drugs and is expressed in the liver. The NAT2 coding region spans 872 base pairs (Genbank Accession No. NM-000015). The PCR product is approximately 1276 base pairs.

Polymorphisms in the NAT2 gene cause the fast and slow N-acetylation phenotypes implicated in the action and toxicity of amine containing drugs. In addition, NAT2 acetylation phenotype is associated with susceptibility to colorectal and bladder cancers. Table 1 summarizes the seven common single nucleotide polymorphisms (SNPs) found in this gene (G191A, C282T, T341C, C481T, G590A, A803G, and G857A) and defines the nine most common alleles (*4 being the wild type allele) along with their associated phenotypes and population frequencies. See Grant et al., *Mutat. Res.* 376:61–70 (1997) and Spielberg et al., *J. Pharmacokint. Biopharm.* 24:509–519 (1996). Each of the seven polymorphisms is a marker for more than one NAT2 allele and each variant allele is defined by two or three SNP substitutions. NAT2 provides a clearly defined, low complexity model system for developing a hybridization based genotyping assay. Typically, homozygous or heterozygous genotypes are made at each polymorphic site before probable allele assignments can be made. In general, individuals who are homozygous for any combination of the slow acetylator alleles are slow acetylators, where rapid acetylators are homozygous or heterozygous for wild-type NAT2 allele. It has been suggested that slow acetylators may be at increased risk for developing bladder, larynx and hepatocellular carcinomas, whereas rapid acetylator may be at risk to develop colorectal cancer. The frequency of the slow acetylator phenotype varies among ethnic groups and is roughly 50%–60% in Caucasian populations. See Grant, D., et al, *Mutation Research* 376:61–70 (1997) and Lin, H., et al, *Pharmacogenetics* 4:125–134 (1994). Polynucleotide array can be used to determine whether a target nucleic acid sequence has one or more nucleotides identical to or different from a specific reference sequence.

TABLE 2

Polymorphisms, alleles, and phenotypes of the NAT2 gene

	Polymor.								
allele	G191A	C282T	T341C	C481T	G590A	A803G	G857A	Phenotype	Frequency
*4	G	C	T	C	G	A	G	Rapid	23.40
*5A	G	C	C	T	G	A	G	Slow	2.50
*5B	G	C	C	T	G	G	G	Slow	40.90
*5C	G	C	C	C	G	G	G	Slow	2.60
*6A	G	T	T	C	A	A	G	Slow	28.40
*7B	G	T	T	C	G	A	A	Slow	2.10
*12A	G	C	T	C	G	G	G	Rapid	0.10
*14A	A	C	T	C	G	A	G	Slow	rare
*14B	A	T	T	C	G	A	G	Slow	0.10
Amino Acid Change	R → Q	None	I → T	None	R → Q	K → R	G → E		

EXAMPLE 9

Performing, Large Numbers of PCR Reactions Using Array-immobilized and Releaseable Primers and Detecting Amplified Product by Hybridization

In the current Example, G 191A polymorphic site of the NAT2 gene was determined using array assembly. A first array with 535 individual derivatized areas (500 um diameter, 1000 um spacing) were prepared on a chip according to previous Examples. Primers and probes are designed and organized as described in FIG. 13 (SEQ. ID. Nos. 6–13). The complementary sequences to the primers are immobilized in the primer sites F and R, to capture respectively Reverse and Forward primers during a pre-hybridization step. Polynucleotide synthesis was carried out in the direction 3' to 5' (3' attached to the surface of the glass) according to known procedures.

A second array was microfabricated according to Example 7 (coverslip). A coverslip contains 10 unit cells (5 mm×5 mm ×100 um depth). Each cell was designed to face 25 derivatized sites on the first array upon juxtaposition of the two arrays.

Pre-hybridization of the Primers

The first array was hybridized in 30 ml of (2×SSC, 0.1% Tween20, and 50 mM $NaHCO_3$) with 2.5 nm of Cyanine 3 Labeled primers at RT for 30 minutes, then washed in 30 mls of 150 mM $NaHCO_3$ at RT for 15 minutes shaking slowly. The first array was then dipped briefly in 70% Ethanol/H_2O and air dried for 10–15 minutes and scanned.

The scanner generated picture on FIG. 14 shows the locations where the labeled primers have accumulated (bright round spots). The high intensity areas correspond exactly to the sites on the surface of the array where the complementary sequences of the primers were synthesized (see FIG. 13). The maximum intensity obtained for this hybridization corresponds to approximatively 30 fmoles of primer hybridized per square millimeter. The level of fluorescence detectable for the other synthesis sites was negligible.

Preparation of Array Assembly

A liquid polymer (100 ul of 2× Self-Seal, MJR) was deposited on the perimeter of the first chip and allowed to dry. Upon assembly and rehydration the liquid polymer created an efficient leakproof gasket that prevented excessive evaporation during the temperature cycling

Array Assembly

The PCR mix (200 ul, 26mM Tris-HCl, pH 9.5, 6.5 mM $MgCl_2$, 0.1% TritonX-100, 20 mM KCl, 200 mM dNTP's, 0.5 U.ul Thermosequenas,) containing 0.1 ng/ul of DNA target (genomic DNA) was used.

The top edge of the first array was put in contact with the second array (coverslip) with a 45 to 60 degrees angle. The PCR mix was loaded in the space created between the first array and the coverslip. The first array was then gradually lowered against the coverslip in a motion that prevents air from being trapped inside the assembly (FIG. 15).

Temperature Cycling

The assembly was introduced into a commercially available in-situ PCR instrument (MJResearch) and the following temperature program was started: 2 min 86° C.; 0.5 min 86° C. ;1.0 min 56.5° C.: 30 times then 5 min 72° C., total cycle time is 1.3 H.

After the last cycle, the assembly is heated to 92° C. for 2 minutes followed by a 25° C. step for 30 min to allow denaturation and hybridization of the PCR product to the probes on the first array.

The assembly was then opened, and the first array was washed with 30 ml of Phosphate Buffer (100 mM NaCl, 3 mM KCl, 4 mM Na_2HPO_4, 1.5 mM NaH_2PO_4, 0.5% Tween20, pH=7.2) for 15 minutes at RT. The first array was then dipped briefly in 70% Ethanol/H_2O and air dried for 10–15 minutes and scanned (FIG. 16).

FIG. 16 is an image of the top four PCR wells of the first array after PCR cycling, hybridization and wash. As before, the primers sites showed very high (gray) or saturating intensities (white). The positive control probe for PCR (PCR positive arrow) showed very high or saturating intensities while the negative control probes (PCR negative arrow) remained close to noise intensity (black). This result demonstrates that a PCR product was specifically generated inside the array assembly.

The Coding Probe 1-A (SNP Detected arrow) had higher intensity than the Coding Probe 2-G (refer to FIG. 13 for exact location). This result agrees with the DNA type (homozygote type 1-A) as determined by dideoxysequencing of the NAT2 PCR product.

EXAMPLE 10

Comparison of Primers Introduced by Hybridization Versus Introduction of the Primers in Solution

Four array assemblies were created as described in Example 9 except for two assemblies where a 200 nM final concentration of the Cyanine 3 labeled primers was added to the PCR mix. The PCR cycle, washes and scans were performed as previously described and the results are summarized in the FIG. 17.

The bars in FIG. 17 represent the average intensities (expressed in Relative Fluorescence Units) collected from the scan for each feature of the polynucleotide-immobilized array.

The introduction of the primers by hybridization gives comparable results to the PCRs with primers added in solution suggesting that the efficiency of the formation of PCR product is similar.

EXAMPLE 11

Performing Large Numbers of PCR Reactions Using Array-immobilized and Releaseable Primers and Detecting Amplified Product by Allele Specific Extension

A DNA microarray with 535 individual derivatized areas (500 um diameter, 1000 um spacing) were prepared according previous Examples. In situ polynucleotide synthesis was carried out in the direction 5' to 3' (5' attached to the surface of the glass) according to known procedures. In this manner the 3' end of the probe is free to be elongated by the DNA polymerase upon hybridization of the amplified products.

A second microfabricated array was prepared according to Example 7 that contains 10 unit cells (5 mm×5 mm×100 um depth). Each cell was designed to face 25 derivatized sites on the DNA microarray upon juxtaposition of the two arrays. Primers and probes were designed and organized as described in FIG. 18 (SEQ. ID. Nos. 14–32).

The preparation of the DNA microarray and application of the liquid polymer is done as described in Example 9.

Reaction Assembly

The PCR mix (200 ul, 26 mM Tris-HCl, pH 9.5, 6.5 mM $MgCl_2$, 0.1%TritonX-100, 20 mM KCl,200 mM dATP, 200 mM dGTP, 175 mM dCTP, 25 mM Cy3-dCTP, 175 mM dTTP, 25 mM Cy3-dUTP, 0.5 U.ul Thermosequenase, 200 nM of each primer) containing 0.1 ng/ul of DNA target (genomic DNA A191G type A homozygote) is used.

Temperature Cycling

The assembly is introduced into a commercially available in-situ PCR instrument (MJR) and the following temperature program is started: 2min 86° C.; 0.5 min 86° C. ;1.0 min 56.5° C.: 30 times then 5 min 72° C., total cycle time is 1.3 H

After completion of the temperature program, the assembly was opened, and the DNA microarray was washed with a solution of triethylamine:ethanol 1:1 for 15 minutes at room temperature, rinsed 3 times with water, dried with nitrogen and scanned. The results are summarized in FIG. 19.

The bars in FIG. 19 represent the average intensities (expressed in Relative Fluorescence Units) collected from the scan for each feature of the DNA microarray.

The Negative Control (C-) intensity corresponds to the background level for this experiment and can be appropriately subtracted for more accurate results. The Probe Coding A (PC(n A) and Probe Non-Coding T (PN(n)T) have higher intensities than respectively the Probe Coding G (PC(n G) and Probe Non-Coding C (PN(n C) in accordance with the DNA type(homozygote type 1-A) as determined by dideoxy-sequencing of the NAT-2 PCR product.

This result shows that the PCR product can be efficiently detected and genotyped using simultaneous PCR, labeling and allele specific extension of the probes at the surface of an array.

EXAMPLE 12

Real-Time Quantitation of Amplification Product

The fluorescent dye SYBR Green I binds to the minor groove of the DNA double helix. In solution, the unbound dye exhibits very little fluorescence, however, fluorescence is greatly enhanced upon DNA-binding. Since SYBR Green I dye is very stable (only 6% of the activity is lost during 30 amplification cycles), it may be used to measure total DNA.

At the beginning of amplification, the reaction mixture contains the denatured DNA, the primers, and the dye. The unbound dye molecules weakly fluoresce, producing a minimal background fluorescence signal which is subtracted during computer analysis. After annealing of the primers, a few dye molecules can bind to the double strand. DNA binding results in a dramatic increase of the SYBR Green I molecules to emit light upon excitation.

During elongation, more and more dye molecules bind to the newly synthesized DNA. If the reaction is monitored continuously, an increase in fluorescence is viewed in real-time. Upon denaturation of the DNA for the next heating cycle, the dye molecules are released and the fluorescence signal falls.

Fluorescence measurement at the end of the elongation step of every PCR cycle is performed to monitor the increasing amount of amplified DNA. Together with a melting curve analysis performed subsequently to the PCR, the SYBR Green I format provides an excellent tool for specific product identification and quantification.

EXEMPLE 13

End-point Quantitation of PCR Product Using Array Assembly

A 10 well coverslip (5 mmx5 mmx100 um) was microfabricated as previously described. A "blank" DNA microarray (surface derivatized, but no DNA synthesis) was used as an inert counterpart for the array assembly.

Reaction Assembly

The PCR mix (200 ul, 26 mM Tris-HCl, pH 9.5, 6.5 mM $MgCl_2$, 0.1% TritonX-100, 20 mM KCl, 200 mM dATP, 200 mM dGTP, 200 mM dCTP, 200 mM dTTP, 0.5 U.ul Thermosequenase, 200 nM of each primer) containing 0.1 ng/ul of DNA target (or no DNA) and 1x Sybr Green was used to prepare the array assembly as described in Example 9.

A picture was taken immediately following the assembly using an apparatus described in FIG. **20**. The array assembly seats on the stage under the CCD Camera. The excitation was generated by a 450W Xenon lamp and filtered using a 485 nm +/−20 nm interference filter. The light was directed towards the surface of the assembly using two optical fibers with a linear shaped oulet to create an homogeneous epi-illumination of the array assembly. The intensity was approxymatively 1.5 uW/cm^2. The fluorescence was collected using a 16-bit, 1024x1024 pixels, back-illuminated CCD camera through a 521 nm +/−20 nm interference filter and a 50 mmx2.8 f lens. Acquisition time was set to 5 secondes, the image generated was displayed on a computer screen for further analysis.

Areas of interest were selected that matched the locations of the PCR wells and the intensities of the pixels comprising the area were averaged.

Temperature Cycling

The assembly is introduced into a commercially available in-situ PCR instrument (MJ Research) and the following temperature program was started: 2 min 86° C.; 0.5 min 86° C. ;1.0 min 56.5° C.: 30 times then 5 min 72° C., total cycle time is 1.3 H

After the completion of the temperature program, the array assembly was imaged as described earlier. The images were analyzed and the results are summarized in FIG. **21**.

The intensity of the fluorescence in the wells containing the template DNA after PCR was higher than before PCR and surpassed the intensity of the wells containing no template before or after PCR. This result shows that the formation of the double stranded PCR product can be monitored and quantitated during the PCR, simultaneously in all of the PCR wells. Accurate monitoring of the increasing amount of amplified DNA provides an excellent tool for gene expression analysis.

EXAMPLE 14

Allele Discrimination Using the 5' Nuclease Assay Using Array Assembly

The 5' nuclease allele discrimination assay exploits the 5'–3' nuclease activity of DNA polymerases to allow direct detection of the PCR product by the release of a fluorescent reporter as a result of PCR. Two probes are used, one probe for each allele in a two-allele system. Each probe consists of a polynucleotide with a 5'-reporter dye and a 3'-quencher dye. TET (6-carboxy-4,7,2',7'-tetrachlorofluorescein) is attached to the 5' end of the probe for the detection of Allele 1. FAM (6-carboxyfluorescein) is attached to the 5' end of the probe for the detection of Allele 2. Each of the reporters is quenched by TAMRA (6-carboxy-N,N,N',N'-tetramethylrhodamine) attached via a linker arm located at the 3' end of each probe.

When the probe is intact, the proximity of the reporter dye to the quencher dye results in suppression of the reporter fluorescence. The probes hybridize to a target sequence within the PCR product. The DNA polymerase cleaves the probe with its 5'–3' nuclease activity. The reporter dye and quencher dye are separated upon cleavage, resulting in increased fluorescence of the reporter. Both primer and probe must hybridize to their targets for amplification and cleavage to occur.

The rate of cleavage of each of the allele specific probes depends directly on the rate of hybridization of the probes.

Detection of the alleles is done by measuring the increase of reporter fluorescence following PCR for each of the probes. The ratio of the intensities of fluorescence is used do determine the polymorphic type of the sample.

2400 individual hydrophilic areas (250 um diameter, 500 um spacing) are prepared on a first array and the solid phase synthesis of the DNA polynucleotides is carried out in the direction 5' to 3' (3' attached to the surface of the glass).

All the synthesis sites are first derivatized with the photocleavable linker to provide a mechanism for all primers and probes to be released into their PCR well A second microfabricated array is prepared according to previous Examples that contains 600 unit cells (1 mm×1 mm×100 um depth), each cell is designed to face four hydrophilic sites on the first array upon juxtaposition of the two arrays.

Primers and probes are designed and organized as described in Table 3.

TABLE 3

Primer Forward Probe Type 2	Probe Type 1 Primer Reverse

While the primers are unmodified DNA sequences amplifying the DNA region comprising the SNP site, allele detection using 5' nuclease assays for real time quantitative PCR requires the two probes to be labeled with distinctly different reporter dyes at the 5' end and a quencher dye at the 3' end of the sequence. The reporter dyes are TET and 6-FAM and the quencher is TAMRA for the 5' and 3' labeling of the sequence. The TET, 6-FAM and TAMRA dyes are commercially available in their phosphoramidite form and are therefore conveniently directly coupled to the growing polynucleotides during in situ synthesis.

Preparation of the DNA Microarray for Assembly

After deprotection of the microarray 100 ul of liquid polymer (2× Self-Seal, MJR) is deposited on the perimeter of the DNA chip and allowed to dry. Upon assembly and rehydratation the liquid polymer will create an efficient leakproof gasket that prevents excessive evaporation during the temperature cycling

Reaction Assembly

200 ul of the PCR mix (26 mM Tris-HCl, pH 9.5, 6.5 mM $MgCl_2$, 0.1%TritonX-100, 20 mM KCl, 200 mM dNTP's, 0.5 U Thermosequenase) containing 0.1 ng/ul of DNA target (genomic DNA) is used to prepare the array PCR assembly. The top edge of the microarray is put in contact with the coverslip with a 45 to 60 degrees angle. The PCR mix is loaded in the space created between the DNA microarray and the coverslip. The DNA chip is then gradually lowered against the coverslip in a motion that prevent air from being trapped inside the assembly.

Cleavage of the Primers and Probes

The assembly is exposed to 365 nm light, measured at 16 mW/cm2, for 15 minutes.

Temperature Cycling

The assembly is introduced into a commercially available in-situ PCR instrument (MJR) and the following temperature program is started: 2 min 86 C.; 0.5 min 86 C. ; 1.0 min 56.5 C.: 30 times then 5.min 72 C., total cycle time is 1.3 H.

After the last cycle, the assembly is heated to 92 C. for 2 minutes followed by a 45 C. step for 30 min to allow denaturation and hybridization of the PCR product to the probes on the DNA microarray.

After the completion of the temperature program, the assembly is imaged with the apparatus described in previous examples. Two images are taken and analyzed for each assembly using sets of filters compatible with TET and FAM spectral properties. The pictures are analyzed and the intensity ratios of the light emitted by each of the dyes are used to determine the allele of the sample for each of the studied SNPs.

The above description is illustrative and not restrictive. Many variations of the invention will become apparent to those of skill in the art upon review of this disclosure. These variations may be applied without departing from the scope of the invention. The scope of the invention should, therefore, be determined not with reference to the above description, but instead should be determined with reference to the appended claims along with their full scope of equivalents.

All publications, patents or web sites are herein incorporated by reference in their entirety to the same extent as if each individual publication patent or web site was specifically and individually indicated to be incorporated by reference in its entirety

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 32

<210> SEQ ID NO 1
<211> LENGTH: 17
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 1

ctcaacccga atctccg 17

<210> SEQ ID NO 2
<211> LENGTH: 15
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 2

gcatgcatgc atgca 15

<210> SEQ ID NO 3
<211> LENGTH: 23
<212> TYPE: DNA

-continued

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 3

ttttatcgga gattcgggtt gag 23

<210> SEQ ID NO 4
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 4

ctcaacccga atctccgata aaa 23

<210> SEQ ID NO 5
<211> LENGTH: 29
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 5

gatgctaccg tgactgactg actgactga 29

<210> SEQ ID NO 6
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 6

ctcaacccga atctccgata aaa 23

<210> SEQ ID NO 7
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 7

cgggtcatgt cttcaactaa ctg 23

<210> SEQ ID NO 8
<211> LENGTH: 15
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 8

ctttggaccc accca 15

<210> SEQ ID NO 9
<211> LENGTH: 13
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 9

ctttgggccc acc 13

<210> SEQ ID NO 10
<211> LENGTH: 17
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 10

tgggtgggtc caaagaa 17

<210> SEQ ID NO 11
<211> LENGTH: 17

-continued

<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 11

tgggtgggcc caaagaa 17

<210> SEQ ID NO 12
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 12

ttttatcgga gattcgggtt gag 23

<210> SEQ ID NO 13
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 13

ggagagagaa gacagttcgt ctt 23

<210> SEQ ID NO 14
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 14

aaaatagcct ctaagcccaa ctc 23

<210> SEQ ID NO 15
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 15

gtcaatcaac ttctgtactg ggc 23

<210> SEQ ID NO 16
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 16

tcacattgta agaagaaacc a 21

<210> SEQ ID NO 17
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 17

tcacattgta agaagaaacc c 21

<210> SEQ ID NO 18
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 18

tcacattgta agaagaaacc g 21

<210> SEQ ID NO 19

-continued

<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 19

tcacattgta agaagaaacc t 21

<210> SEQ ID NO 20
<211> LENGTH: 17
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 20

gagacaccac ccaccca 17

<210> SEQ ID NO 21
<211> LENGTH: 17
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 21

gagacaccac ccacccc 17

<210> SEQ ID NO 22
<211> LENGTH: 17
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 22

gagacaccac ccacccg 17

<210> SEQ ID NO 23
<211> LENGTH: 17
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 23

gagacaccac ccaccct 17

<210> SEQ ID NO 24
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 24

tgatcacatt gtaagaagaa aca 23

<210> SEQ ID NO 25
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 25

tgatcacatt gtaagaagaa acc 23

<210> SEQ ID NO 26
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 26

tgatcacatt gtaagaagaa acg 23

-continued

<210> SEQ ID NO 27
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 27

tgatcacatt gtaagaagaa act 23

<210> SEQ ID NO 28
<211> LENGTH: 17
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 28

ggagacacca cccacca 17

<210> SEQ ID NO 29
<211> LENGTH: 17
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 29

ggagacacca cccaccc 17

<210> SEQ ID NO 30
<211> LENGTH: 17
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 30

ggagacacca cccaccg 17

<210> SEQ ID NO 31
<211> LENGTH: 17
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 31

ggagacacca cccacct 17

<210> SEQ ID NO 32
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 32

actgactgac tgactgactg 20

We claim:

1. A method for performing a plurality of enzymatic amplification reactions, comprising the steps of:

(a) obtaining a first solid support wherein at least two releasable primers of each reaction are confined to areas on the surface of said first solid support;

(b) obtaining a second solid support with a plurality of wells providing separation of said plurality of reactions;

(c) assembling said first and second solid support;

(d) filling said wells with reactants of said reactions;

(e) releasing said releasable polynucleotides into solution in each well; and

(f) performing said plurality of amplification reactions in parallel.

2. The method according to claim 1 wherein said reactants in step (d) comprise a DNA polymerase, a target nucleic acid, and deoxynucleotides.

3. A method for performing a plurality of polynucleotide amplification reactions, comprising the steps of:

(a) obtaining first solid support wherein a plurality of immobilized moieties are confined to a plurality of areas on the surface of said first solid support and each said immobilized moiety contains a releasable site and a primer;

(b) obtaining a second solid support wherein the surface of said second solid support contains a plurality of areas and reactants of said polynucleotide amplification reactions are confined on each area of said second solid support;

(c) assembling said first and second solid supports, wherein said reactants of said polynucleotide amplification reactions on said second solid support are in contact with said immobilized moieties on said first solid support; and

(d) releasing said primers;

(e) performing said plurality of polynucleotide amplification reactions.

4. A method for performing a plurality of polynucleotide amplification reactions and capturing amplification products, comprising the steps of:

(a) obtaining a first solid support wherein a plurality of immobilized moieties are confined to a plurality of areas on the surface of said first solid support and each said immobilized moiety contains a releasable site and a primer;

(b) obtaining a second solid support wherein the surface of said second solid support contains a plurality of areas and reactants of said polynucleotide amplification reactions are confined on each area of said second solid support;

(c) assembling said first and second solid supports, wherein said reactants of said polynucleotide amplification reactions on said second solid support are in contact with said immobilized moieties on said first solid support;

(d) releasing said primers;

(e) generating amplification products of said polynucleotide amplification reactions; and

(f) capturing said amplified products by a plurality of immobilized polynucleotide probes on either said first or second solid support through hybridization.

5. A method for detecting a plurality of polynucleotide sequence variations, comprising the steps of:

(a) obtaining a first solid support wherein a plurality of immobilized moieties are confined to a plurality of areas on the surface of said first solid support and each said immobilized moiety contains a releasable site and a primer;

(b) obtaining a second solid support wherein the surface of said second solid support contains a plurality of areas and reactants of said polynucleotide amplification reactions are confined on each area of said second solid support;

(c) assembling said first and second solid support, wherein said reactants of said polynucleotide amplification reactions on said second solid support are in contact with said immobilized moieties on said first solid support;

(d) releasing said primers;

(e) generating amplification products of said polynucleotide amplification reactions;

(f) capturing said amplified products by a plurality of immobilized polynucleotide probes on either said first or second array through hybridization; and

(g) detecting polynucleotide sequence variations by hybridization complexes in step (f).

6. A method for quantitating polynucleotides in a target nucleic acid, comprising the steps of:

(a) obtaining a first solid support wherein a plurality of immobilized moieties are confined to a plurality of finite areas on the surface of said first solid support and each said immobilized moiety contains a releasable site and a primer;

(b) obtaining a second solid support wherein the surface of said second solid support contains a plurality of areas and reactants of said polynucleotide amplification reactions are confined on each area of said second solid support;

(c) assembling said first and second solid support, wherein said reactants of said polynucleotide amplification reactions on said second solid support are in contact with said immobilized moieties on said first solid support;

(d) releasing said primers;

(e) generating amplification products of said polynucleotide amplification reactions; and

(f) quantitating amplified products.

7. A method for detecting polynucleotide sequence variations in a target nucleic acid, comprising the steps of:

(a) obtaining a first solid support wherein a plurality of immobilized moieties are confined to a plurality of areas on the surface of said first solid support and each said immobilized moiety contains a releasable site and a primer;

(b) obtaining a second solid support wherein the surface of said second solid support contains a plurality of areas and reactants of said polynucleotide amplification reactions are confined on each areas of said second solid support;

(c) assembling said first and second solid supports, wherein said reactants of said polynucleotide amplification reactions on said second solid support are in contact with said immobilized moieties on said first solid support;

(d) releasing said primers;

(e) generating amplification products of said polynucleotide amplification reactions; and

(f) capturing said amplified products by a plurality of immobilized polynucleotide probes on either first or second array though hybridization; and

(g) detecting polynucleotide sequence variations by a polynucleotide modifying enzyme.

8. A method for amplifying a target nucleic acid, capturing the amplified product, and detecting a polynucleotide sequence variation in the amplified product, comprising the steps of:

(a) obtaining a first solid support wherein:

(1) the surface of said first solid support comprises a first, second, third, and fourth areas;

(2) a first chemical moiety, comprising a releasable forward primer for said target nucleic acid, is immobilized on said first area;

(3) a second chemical moiety, comprising a releasable reverse primer for said target nucleic acid, is immobilized on said second area;

(4) a first polynucleotide probe, comprising a subsequence complementary to one variant of said polynucleotide variation, is immobilized on said third area said subsequence containing at least one interrogation position complementary to a corresponding nucleotide in said variant; and

(5) a second polynucleotide probe is immobilized to said fourth area, said second polynucleotide probe differing from said first polynucleotide probe by at least one nucleotide;

(b) obtaining a second solid support wherein the surface of said solid support comprises a reaction well and a mixture of reacts comprising a DNA polymerase, said

target nucleic acid, and deoxynucleotides are placed within said reaction well;

(c) assembling said first and second solid support, wherein said mixture of reactants are in contact with said first, second, third, and fourth areas on said first solid support;

(d) releasing said releasable forward and reverse primers;

(e) generating the amplified product for said target nucleic acid;

(f) capturing the amplified product by said first or second polynucleotide probe through hybridization;

(g) washing said first solid support;

(h) comparing the relative binding of two probes on said first solid support; and

(i) identifying said polynucleotide variation in the amplified product.

9. A method for amplifying a target nucleic acid, capturing the amplified product, and detecting a polynucleotide sequence variation in the amplified product, comprising the steps of:

(a) obtaining a first solid support wherein:

(1) the surface of said first array comprises a first, second, third, and fourth areas;

(2) a first chemical moiety, comprising a releasable forward primer specific for said region of said target nucleic acid, is immobilized on said first area;

(3) a second chemical moiety, comprising a releasable reverse primer specific for said region of said target nucleic acid, is immobilized on said second area;

(4) a first polynucleotide probe, comprising a subsequence complementary to one variant of said polynucleotide variation, is immobilized on said third area, said subsequence containing at least one interrogation position complementary to a corresponding nucleotide in said variant; and

(5) a second polynucleotide probe is immobilized to said fourth area, said second probe differing from said first probe by at least one nucleotide;

(b) obtaining a second solid support wherein the surface of said solid support comprises a reaction well and a mixture of reactants comprising a DNA polymerase, said target nucleic acid, and deoxynucleotides are placed within said reaction well;

(c) assembling said first and second solid support, wherein said mixture of reactants are in contact with said first, second, third, and fourth areas on said first solid support;

(d) releasing said releasable forward and reverse primers;

(e) generating the amplified product for said target nucleic acid;

(f) capturing the amplified product by said first or second polynucleotide probes through hybridization;

(g) extending said one or more hybridization complexes in step (f);

(h) washing said first solid support; and

(i) identifying said polynucleotide variation using said one or more extended products in step (g).

10. The method according to any one of the claims 1–2 and 3–9 wherein said areas on the surface of said first solid support are hydrophilic areas.

11. The method according to any one of the claims 1–2 and 3–9 wherein the density of areas on said first solid support is between about 10 to 10,000 per cm^2.

12. The method according to any one of the claims 1–2 and 3–9 wherein the size of each said area on said first solid support is between about 10^{-3} to 5 mm^2.

13. The method according to any one of the claims 1–2 and 3–9 wherein the number of areas on said first solid support is between about 10 to 500,000.

14. The method according to any one of the claims 3–9 wherein said immobilized moieties are prepared by in situ synthesis.

15. The method according to any one of the claims 3–9 wherein said immobilized moieties are prepared by spotting.

16. The method according to any one of the claims 3–9 wherein said immobilized moieties are prepared using an ink-jet printing like device.

17. The method according to any one of the claims 3–9 wherein each of said plurality of areas on said second solid support is a reaction well.

18. The method according to any one of the claims 3–9 wherein said immobilized moiety on each area of said first solid support in step (a) is a hybridization complex between an immobilized polynucleotide hybridized with a releasable primer for said polynucleotide amplification reactions.

19. The method according to any one of the claims 3–9 wherein said immobilized moiety on each area of said first solid support in step (a) comprises a cleavable site and a releasable primer.

20. The method according to any one of the claims 3–9 wherein said immobilized moiety on each area of said first solid support in step (a) comprises a cleavable site and a releasable primer for said polynucleotide amplification reactions and said cleavable site is cleavable by photolysis.

21. The method according to any one of the claims 3–9 wherein said immobilized moiety on each area of said first solid support comprises a cleavable site and a releasable primer for said polynucleotide amplification reactions and said cleavable site comprises an o-nitrobenzyl linker.

22. The method according to any one of the claims 1–2 and 3–9 wherein said first or second solid support is glass.

23. The method according to claim 1 wherein said releasable primers are released.by photolysis in step (e).

24. The method according to claim 1 wherein said releasable primers are released by strand separation in step (e).

25. The method according to claim 1 wherein said releasable primers are released by enzymatic reaction in step (e).

* * * * *